
# GEOProcessing 2013

The Fifth International Conference on Advanced Geographic Information Systems, Applications, and Services

ISBN: 978-1-61208-251-6

February 24 - March 1, 2013

Nice, France

**GEOProcessing 2013 Editors**

Claus-Peter Rückemann, Leibniz Universität Hannover / Westfälische Wilhelms-Universität Münster / North-German Supercomputing Alliance (HLRN), Germany

# GEOProcessing 2013

# Forward

The fifth edition of The International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2013), held in Nice, France, February 24 - March 1, 2013, addressed the aspects of managing geographical information and web services.

The goal of the GEOProcessing 2013 conference was to bring together researchers from the academia and practitioners from the industry in order to address fundamentals of advances in geographic information systems and the new applications related to them using the Web Services. Such systems can be used for assessment, modeling and prognosis of emergencies

GEOProcessing 2013 provided a forum where researchers were able to present recent research results and new research problems and directions related to them. The topics covered aspects from fundamentals to more specialized topics such as 2D & 3D information visualization, web services and geospatial systems, geoinformation processing, and spatial data infrastructure.

We take this opportunity to thank all the members of the GEOProcessing 2013 Technical Program Committee as well as the numerous reviewers. The creation of such a broad and high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to the GEOProcessing 2013. We truly believe that, thanks to all these efforts, the final conference program consists of top quality contributions.

This event could also not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the GEOProcessing 2013 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that GEOProcessing 2013 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in geographic information research.

We also hope that Côte d'Azur provided a pleasant environment during the conference and everyone saved some time for exploring the Mediterranean Coast.

**GEOProcessing 2013 Chairs**

Monica De Martino, Consiglio Nazionale delle Ricerche - Genova, Italy

Claus-Peter Rückemann, Leibniz Universität Hannover / Westfälische Wilhelms-Universität Münster / North-German Supercomputing Alliance (HLRN), Germany
Bernd Resch, Massachusetts Institute of Technology - Cambridge, USA

# GEOProcessing 2013

# Committee

**GEOProcessing Advisory Chairs**

Monica De Martino, Consiglio Nazionale delle Ricerche - Genova, Italy
Claus-Peter Rückemann, Leibniz Universität Hannover / Westfälische Wilhelms-Universität Münster / North-German Supercomputing Alliance (HLRN), Germany
Bernd Resch, Massachusetts Institute of Technology - Cambridge, USA

**GEOProcessing 2013 Technical Program Committee**

Diana F. Adamatti, Universidade Federal do Rio Grande, Brazil
Riccado Albertoni, Universidad Politécnica de Madrid, Spain
Francesc Antón Castro, Denmark´s National Space Institute, Denmark
Thierry Badard, Université Laval - Québec, Canada
Petko Bakalov, Environmental Systems Research Institute, USA
Fabian D. Barbato, ORT University - Montevideo, Uruguay
Thomas Barkowsky, University of Bremen, Germany
Reinaldo Bezerra Braga, Joseph Fourier University - Grenoble 1, France
Budhendra L. Bhaduri, Oak Ridge National Laboratory, USA
Ling Bian, University at Buffalo, USA
Sandro Bimonte, Irstea | TSCF - Clermont Ferrand, France
Giuseppe Borruso, University of Trieste, Italy
Boyan Brodaric, Geological Survey of Canada, Canada
Mete Celik, Erciyes University, Turkey
Xin Chen, NAVTEQ Corporation - Chicago, USA
Chi-Yin Chow, City University of Hong Kong, Hong Kong
Christophe Claramunt Naval Academy Research Institute, France
Eliseo Clementini, University of L'Aquila, Italy
Ana Cristina Costa, Universidade Nova de Lisboa, Portugal
Alfredo Cuzzocrea, ICAR-CNR & University of Calabria, Italy
Chenyun Dai, Purdue University, USA
Monica De Martino, Consiglio Nazionale delle Ricerche (CNR) - Genova, Italy
Anselmo C. de Paiva, Universidade Federal do Maranhão, Brazil
Cláudio de Souza Baptista, University of Campina Grande, Brazil
Antonios Deligiannakis, Technical University of Crete, Greece
Yerach Doytsher, Technion - Israel Institute of Technology, Haifa, Israel
Suzana Dragicevic, Simon Fraser University- Burnaby, Canada
Javier Estornell Cremades, Universidad Politecnica de Valencia, Spain
Aly A. Farag, University of Louisville, USA
Michael P. Finn, Center of Excellence for Geospatial Information Science (CEGIS), U.S. Geological Survey, Denver, USA
Stewart Fotheringham, University of St Andrews in Scotland, UK

**Copyright Information**

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

**Table of Contents**

# A Service-Oriented Platform to Map Cycling Routes

George Henrique Rangel Costa
Department of Computer Science
Santa Catarina State University
Joinville, Brazil
dcc6ghrc@joinville.udesc.br

Fabiano Baldo
Department of Computer Science
Santa Catarina State University
Joinville, Brazil
baldo@joinville.udesc.br

Leandro Lopes Taveira
Department of Computer Science
Santa Catarina State University
Joinville, Brazil
leandrolt@gmail.com

***Abstract*—Over the last few years bicycles have been gaining more attention than motorized vehicles due to the benefits that they offer. Both private and public sectors are joining efforts to stimulate its usage, and an important step in this process is to bring cycling information closer to the population by making it easily available. Computer systems can be used to mitigate such gap by providing information like cycle routes, bicycle stop places and so forth. Considering current technologies it is mandatory that not only personal computers but also smartphones and other mobile devices should be able to access such information. However, to deal with this range of devices it is necessary to use technologies that support interoperability. In this context, a platform using service-oriented architecture and standard communication protocols can wrap a geographic database and a map server to transparently support the presentation of geographic and non-geographic information related to cycling. This paper presents the design and development of such a solution, which aims to increase the availability of cycling information and reduce the coupling between its storage and presentation. To assess this approach, a test scenario is built using real data and two applications are developed – one for Android smartphones and another one for web browsers.**

***Index Terms*—web mapping; web service; web map service; cycling routes.**

## I. Introduction

In order to move inside the cities in a faster way people are buying more cars and motorcycles. However, the direct consequences of this attitude are the increase of traffic jam and air pollution [14]. In this scenario, bicycles represent a healthier and environmentally beneficial option. The expansion of cycling routes is an important action to encourage people to use such transport, but this action should not be taken alone. Initiatives to increase people's awareness regarding the existence and location of cycling routes are also necessary and indirectly contribute to increase the number of people using bicycles.

Computer systems can contribute to these initiatives since they facilitate manipulation of maps and related data, i.e., geographic and non-geographic information, respectively. Moreover, this information should be available to various types of electronic devices, such as smartphones and personal computers, in order to disseminate cycling information to all interested people no matter where they may be. Therefore, a problem to be solved is how to facilitate access to cycling information in order to increase availability and reduce the coupling between its storage and presentation.

Several technologies can be used, but none of them are, individually, a complete solution. For example, map servers answer the question of how to provide maps with cycling routes. However, it is also necessary to access non-geographic information, as the route's description and photos of points of interest, which map servers commonly do not support.

Another point that should be considered is the interoperability, which is essential to distribute information to different devices. In the last years, one of the most prominent technologies to enable interoperation is the Service-Oriented Architecture (SOA). One of the greatest advantages of SOA is that it uses standard communication technologies. A platform based on this architecture could answer client applications requests in a standardized way, transparently obtaining geographic information from a map server and non-geographic information from a relational database.

Based on these assumptions, the objective of this work is to design and develop a service-oriented platform that provides cycling information to electronic devices.

The methodology that drives this work begins with a questionnaire applied to groups of cyclists from Joinville-Brazil. This questionnaire aims to identify the difficulties faced by cyclists in accessing cycling information. The analysis of the collected answers is the basis for defining the platform requirements. Then, a literature review of relevant concepts and technologies is done, as well as the analysis of related work. After collecting the needed assets, the platform is designed and implemented. Finally, a test scenario is built and verification and validation tests are proposed in order to evaluate the results of the proposed approach. These tests are executed in an application for smartphones with Android operating system and in an application for web browsers.

The paper is structured as follows. Section 2 presents a brief review of the technologies used to exchange geographic information. Section 3 presents other initiatives for providing geographic and non-geographic data on the web. Section 4 details the development of the service-oriented platform. Section 5 explains how the verification and validation tests were performed. Finally, conclusion and possible future work are drawn.

 ISBN: 978-1-61208-251-6

## II. Technologies for Geographic Data Interchange

Geographic Information Systems (GIS) are special types of software devoted to store, analyze, and display geographic information in digital format [11]. There are different types of GIS, each one applied to support a subset of functionalities related to geographic data manipulation. One example of specialized GIS ia a map server. It can be considered a GIS with limited capabilities, because it only provides the means to visualize, on the Internet, geographic information usually stored in geographic databases [12] [16].

Map servers have a client-server architecture where at the client side there is a web mapping application that runs on a web browser. Such client applications use an Application Programming Interface (API) where the users make requests, commonly asynchronous, for images that correspond to parts of a map. The map server receives these requests, generates the images and sends them back to the web mapping application [5] [17].

In order to establish proper communication between client and server it is necessary that both use the same protocol. In this sense, Web Map Service (WMS) is a standard for geographic information retrieval in either vector or graphic formats. Its main operations are GetMap and GetFeatureInfo, which return the map and additional information about a point, respectively. There is also the GetCapabilities operation, which provides meta-data about the service, such as the parameters accepted by the other operations. The WMS operations are called through requests in Uniform Resource Locator (URL) format, which simplifies the adoption of this standard for web mapping applications [18].

## III. Initiatives for Providing Cycling Geographic Information on the Web

As the number of people using smartphones and other devices with access to the Global Positioning System (GPS) grows, web mapping applications become increasingly popular and every day new ones are developed. Many of these new applications are devoted to stimulate the usage of bicycles and try to address problems faced by cyclists.

Some applications offer features such as route planning. CycleStreets [4], whose target audience are people that live in England, is an example. Given the desired origin, destination and speed the system provides route options ranging from ”fastest” to ”quietest”, considering shortcuts and lanes exclusive for cyclists. It also shows useful non-geographic information, such as the number of traffic lights along the route. In addition to planning routes, users can view photos and descriptions of points of interest. The application has implementations for web browser and smartphones with Android or iOS operating systems, which facilitates and encourages its use. Developers who want to use the services offered by CycleStreets should use an API that communicates with their server using GET requests [3]. However, the maps must be obtained from an external source, such as OpenStreetMap (OSM) [13].

Another application is called Biketastic [2] and aims to facilitate sharing of cycling routes with cyclists in the region of Los Angeles-USA. Using a smartphone application, cyclists can record their path while riding a bicycle. Besides the geographic information, the application may use the device’s microphone to capture the environment noise level and the accelerometer to identify possible holes or other defects on the path’s surface. Users can add photos and videos of points of interest if they wish. At the end of the ride all this information is sent to a server in order to become available to other users, which can access it using either a smartphone or a web browser [15]. This application does not provide ways for other developers to access its cycling information.

According to the issues presented above, it is clear that CycleStreets and Biketastic contribute to promote the usage of bicycles in the regions where they are offered. Therefore, it is important to develop similar solutions for other regions and countries. However, some points regarding both works should be considered. CycleStreets does not display the points of interest registered by users on top of calculated routes; a feature that could help users choose the desired route. On the other hand, Biketastic sorts the routes by the month of submission; a feature that makes it more difficult to find routes in specific regions.

Also, both works do not provide a search engine to find routes based on characteristics such as surface type, lack of proper route signaling, crossed districts, difficulty and so forth. Moreover, they do not differentiate urban cycling routes from rural trails. For the purpose of this work, the designed platform should provide these features. This work aims to stimulate the usage of bicycle not only for work but also for fun. Therefore, users should be capable of searching for cycling routes inside the city, as well as for long distance trails to enjoy the countryside.

Finally, only CycleStreets provides an API for developers. However, it does not provide access to all information needed to build a client application, like the maps. This shortcoming forces that every developed client must connect to another source of data. In order to avoid this, the proposed platform should provide all information that a client application needs to display cycling routes, including the maps.

## IV. Approach for Providing Cycling Information

Besides being an alternative to motorized vehicles, bicycles can also be used for purposes such as have fun, get fit, competitions and so on. For each purpose, different geographic and non-geographic cycling information become useful and all this information should be offered by the proposed platform, such as the route’s name, length, surface type, districts that it crosses, photos of points of interest and so on.

On the other hand, one inherent requirement of this work is that various electronic devices should be able to access the data. Recently, smartphones have gained a lot of attention. These devices support different wireless technologies for Internet access. However, the cost to connect to the Internet using 2G and 3G networks and the connectivity problems outside

 ISBN: 978-1-61208-251-6 

the cities' downtown are constraints to be overcome in order to provide a platform with constant and complete access.

Due to these issues, the first step is to understand what are the users' needs. To do so, a questionnaire was built and put available on the Internet (the online version of the questionnaire is no longer available, but the questions are listed, in Portuguese, in [6]). This questionnaire comprises questions related to data and connection availability, as well as questions about geographic and non-geographic information needed by the several kinds of users and about pervasive computing (useful on smartphone access).

Approximately 300 cyclists were invited to participate and 84 have filled out the questionnaire. By analyzing their answers it was possible to confirm their interest in accessing the platform through smartphones, especially with those devices that use Android (preferred by 39.29% of the cyclists) and iOS operating system (preferred by 38.10% of the cyclists). Despite both operating systems being almost equally requested, Android allows applications to be tested on real devices for free, and the SDK is available for Windows, Linux and MAC OS X [9]. Meanwhile, iOS requires a paid subscription to allow applications to be tested on real devices, and the SDK is available only for MAC OS X [1]. Android is also the most used operating system, with more than 60% of the market share [7]. For these reasons, only the application for Android operating system has been implemented.

The questionnaire also shows users' interest on using the sensors present in today's smartphones to show their current speed and location. Furthermore, 96.43% of the cyclists desire to record their paths, and 95.24% would like to send the recorded information to be stored in the platform in order to enrich its database.

Internet connectivity seems to be an obstacle since 97.65% of the cyclists are interested in copying the information provided by the platform to their own devices in order to avoid constant Internet connection. Only 25.24% of the cyclists want to have offline access to the photos of points of interest, so this functionality will not be implemented.

Considering all the questionnaire results it was possible to realize that the cyclists want to use their mobile devices mainly to access data about the countryside near the city where they live, so as to learn the routes and ride for fun. Meanwhile, as observed in the analysis of related work, the platform should allow developers of client applications to implement functionalities such as search for routes according to a set of predefined characteristics and differentiate between urban and rural routes. Also, it should provide all information needed for a client application to display cycling routes in a map.

Based on the literature review and on the questionnaire assessment now it is possible to define the system requirements.

### A. Requirements

Regarding functional requirements, first the platform should allow client applications to search routes according to characteristics such its type (rural or urban), surface type or lack of proper signage. Also, when a user selects a route its geographic and non-geographic information must be provided to the client application, as well as all photos associated with its points of interests. Finally, the platform should enable the client applications to copy all information regarding a specific route (except the photos, as explained in the previous section) to the electronic device. Doing this, the client device does not need to maintain any external connection.

Regarding the platform's non-functional requirements, it should be accessed either via personal computers – for example, by a web mapping application – or via other types of electronic devices. Also, it should provide all information through a single interface. Finally, the platform must use standardized communication technologies.

These functional and non-functional requirements are the basis for the specification of 11 use cases that guide the project of all elements that comprise the platform.

### B. Service Interface

First, it is important to mention that the platform was designed using the service-oriented approach. Therefore, the requirements were translated into service operations, including their input and output parameters. This approach was used in order to provide clients with a single facade to access all information needed to display maps and cycling routes. So, the only way that a client application can access the platform is through its interface provided as a web service. The platform's web service interface is shown in Figure 1. Among the 13 operations listed, those that support displaying geographic and non-geographic information are *getMap*, *getOSMMap*, *getImage*, *getNonGeographic* and *getFeatureInfo*.

The *getMap* operation returns an image containing the routes and points of interest that exist in a given geographic area. However, these routes should be displayed on top of a map or satellite photo. This data can be obtained from external servers such as OpenStreetMap and Google Maps. In order to simplify the clients' implementation the platform transparently provides maps to the clients using the OpenStreetMap, since its license of use allows this [19]. This feature is implemented through the *getOSMMap* operation. When a client make a call it receives a single image containing the OpenStreetMap's map with the routes already overlaid. Client applications are still free to implement their own access to other external map servers. However, they should use the *getMap* operation and manually overlay the images. It should be noted that the images from Google Maps are not provided through the platform as with OpenStreetMap because its license of use does not allow it [10].

The non-geographic information is obtained using operations like *getImage* – that returns any image related to a route or point of interest – and *getNonGeographic* – that returns all relevant non-geographic information regarding a route, such as difficulty level and description. Finally, the *getFeatureInfo* operation outputs the information about which cycling routes and points of interest exist in a given geographic coordinate.

In order to support routes search, a set of operations whose name starts with "get" were designed, such as *getDistricts* and

 ISBN: 978-1-61208-251-6 

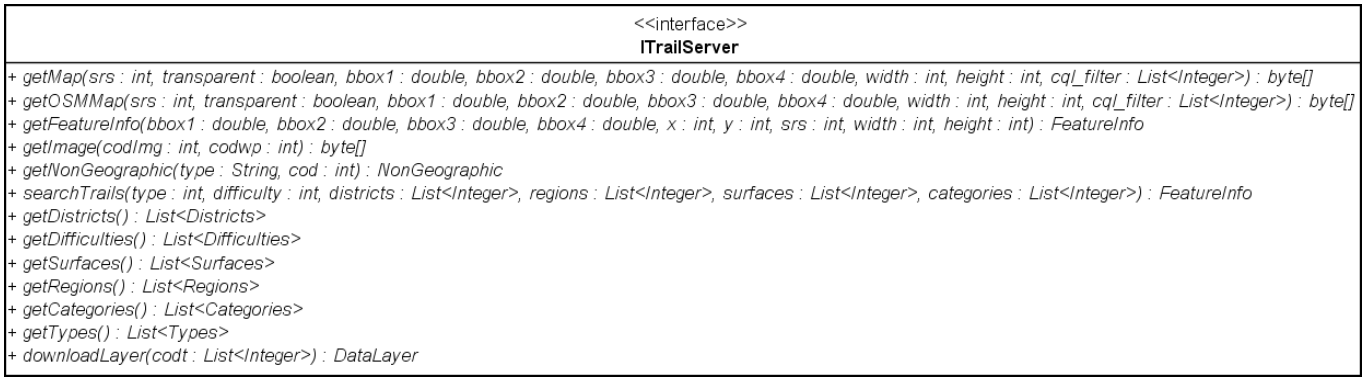



Fig. 1. Web Service Interface.

*getTypes*. These ”get” operations return lists of values that are used to fill in lists of search options. Based on these lists the client application can then call the operation *searchTrails*, which returns a list with all routes that fit the selected search options.

Lastly, the operation that supports the requirement of copying a route to the mobile device is called *downloadLayer*. It returns a list with all geographic and non-geographic information related to the selected route, including its points of interest. This data can then be locally stored in a database at the electronic device, and can be displayed even without Internet connection.

### C. Service Architecture

The operations listed in the web service interface are implemented in a three-layer architecture – business, data and persistence – which can be seen in Figure 2. This is a reduced version of the entire class diagram that highlights where each operation presented in the interface is implemented, as well as important classes and packages. It is important to note that the presentation layer is client dependent, so it is out of the platform's scope. Client applications are represented in the diagram by the *Client* package.

The web service interface shown in Figure 1 is placed in the business layer. Although this layer contains few classes, it processes all the client requests. The access to OpenStreetMap (represented in the diagram by the *OSM* package) using the *getOSMMap* operation is also implemented on it. As explained in the previous section, client applications can connect to other services – such as Google Maps – by themselves, and this is represented in the diagram by the *GMaps* package, connected to *Client*.

The data layer implements all the data objects used by the platform, as well as all operations that manipulate cycling information. The persistence layer manages the database access. Finally, the database reflects all these classes and their relationships in a relational schema.

### D. Implementation

The platform shown in Figure 2 was implemented using the Java programming language in the Netbeans Integrated Development Environment (IDE). The cycling data was stored in the PostgreSQL Database System, which supports geographic data through the extension PostGIS. The JPA library was used to persist data in the database. GeoServer was chosen as the map server since it supports PostgresSQL and implements the WMS standard [8]. Both GeoServer and the platform must run in an application server and Tomcat was chosen for that.

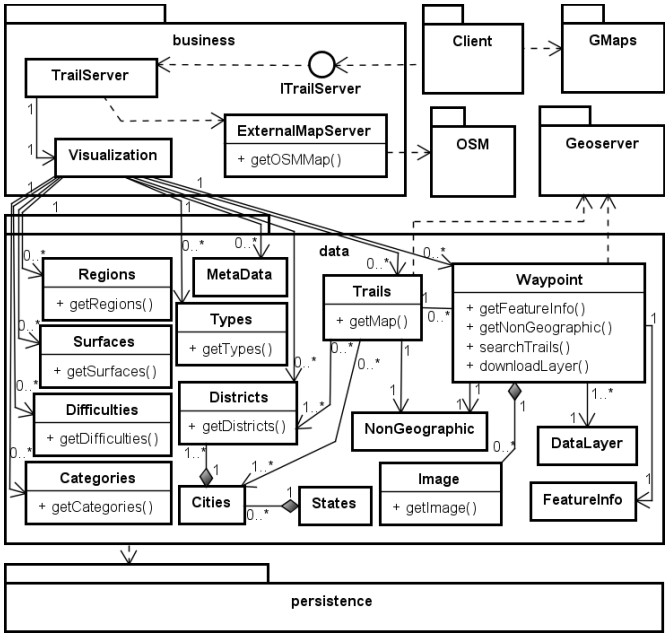



Fig. 2. Platform's Structural Project.

Regarding the implementation, some aspects should be considered. The output of the *getMap*, *getOSMMap* and *getImage* operations are images. However, in SOAP all communications are made through envelopes written in XML format. The SOAP standard has an extension to binary data that uses additional structures to represent the image, but this makes the message even bigger [21]. To reduce the amount of data sent over the network for each request, the adopted solution was to convert the images into byte arrays using Base64. This array is then written in the XML envelope like any other textual data.

Nevertheless, this part of the message needs to be decoded by the client application.

It is also important to mention that both *getMap* and *getFeatureInfo* operations pass clients' requests to the map server as WMS requests. Although it is possible to enable clients to direct access the map server through WMS operations, here it is not allowed because this platform intends to provide a single and centralized interface where client applications can have access to all needed data. This approach facilitates the development of client applications because it hides unnecessary technical issues.

## V. Result Assessment

The methodology that guides the platform's evaluation is composed of verification and validation tests that cover all operations. The verification test used the black box approach, where the requirements compliance is verified by comparing the expected output with the real output. On the other hand, the validation test intended to demonstrate that the platform fulfills its users' expectations, i.e., that it works as expected in everyday situations.

To apply this methodology a test scenario was built using information collected by a group of students involved in the project. They mapped, using GPS devices, a set of urban and rural cycling routes from the city of Joinville. In total, 56 routes and 271 points of interest were recorded. At the end of the collecting process all geographic and non-geographic information was stored in the database designed for this project.

For the verification test a tool called SoapUI was used to create SOAP envelopes and send them to the platform, as well as analyze the SOAP envelopes received back. In each of the 14 test cases the contents of the received envelopes were compared to the information stored in the database and everything was 100% compatible.

For the validation test it was necessary to ensure that different types of devices are capable to access the platform. Thus, two client applications were developed, one for smartphones with the Android operating system and another one for web browsers. The next sections describe the development of each client application, as well as present their test cases.

### A. Smartphone Application

The decision of using the Android operating system is based on the questionnaire results, as explained in Section 4. The questionnaire also justifies the platform's requirement to allow the copy of a route to the client device. As a way to test if the platform correctly implemented this non-functional requirement – the *downloadLayer* operation – the smartphone application was divided into two working modes: online and offline. Both modes have access to the same functionalities. The only difference is in the source of the cycling information. In online mode all data is obtained from the platform, requiring an active Internet connection. In offline mode only the information already stored in the smartphone is used, not requiring an active Internet connection. Users can use


Fig. 3. Smartphone application accessing the platform.

the application's menu to switch the working mode whenever they want. Note that before viewing a route in offline mode the users must switch to online mode and use the copying functionality, which will connect to the platform and request all data related to the selected route. This data will be then saved in the smartphone's database and can be visualized when in offline mode. Also note that these two working modes were implemented only in the smartphone application to serve as an example of how the *downloadLayer* operation could be used by a client.

A total of 30 test cases have been created for the validation test. Figure 3 presents one of these test cases where the client application is in online mode. In Figure 3.a the output from operation *getMap* is shown. When the user touches the map with a finger the application calls the *getFeatureInfo* operation and the balloon in Figure 3.a indicates that this operation is being executed. Figure 3.b shows the result of the *getFeatureInfo* operation. When the user selects one of those routes its non-geographic information is requested using operations such as *getNonGeographic*, *getType* and *getDistrict*, which are presented in Figure 3.c.

This example demonstrates the use of 77% of the operations defined in the platform's interface. Due to space restrictions the other test cases are not presented here, but their results show that all information – including both text and images (binary data) – is properly transferred to the device and displayed on its screen.

### B. Web Browser Application

The web browser client was developed in order to enable users to access the platform through personal computers. However, since personal computers are usually used at home, many features – especially the pervasive ones – would not make sense if implemented, such as show the users' coordinates or alert them about the proximity of points of interest. This client has been implemented as a web mapping application and its purpose is to take advantage from personal computers' screen size and thus better present cycling routes on top of maps or satellite images. Its features include search for cycling routes based on their characteristics and display information about routes and points of interest. This client uses HTML and

 ISBN: 978-1-61208-251-6

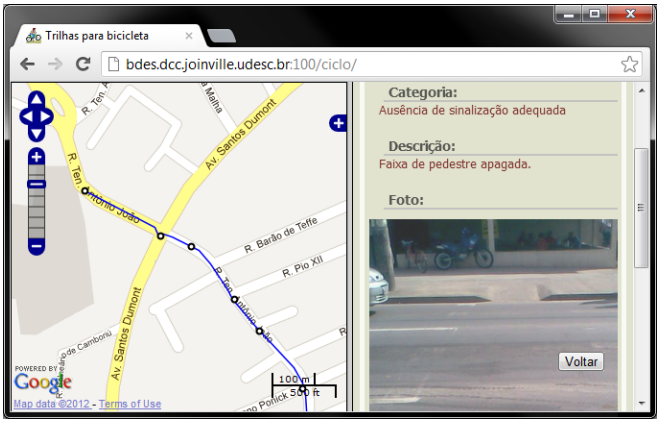



Fig. 4. Web browser accessing the platform.

PHP languages, and the map manipulation is controlled by the OpenLayers library. It can be accessed at [20], and Figure 4 exemplifies its use.

For the validation test 12 test cases were created. The different number of test cases when compared to the smartphone application is explained by the features that could not be implemented in the browser. In Figure 4 the non-geographic information about a point of interest is presented on the right side and the geographic information on the left. The test case in Figure 4 illustrates some operations not used in the smartphone test case shown in Figure 3, such as *getImage* and *getCategory*.

## VI. Conclusion and Future Work

Bicycles are healthier than motorized vehicles and therefore its use should be encouraged. Initiatives like Biketastic and CycleStreets contribute in this direction, but limitations were identified in both works. In order to solve them, this paper presented a service-oriented approach to disseminate information about cycling routes in a device independent way.

The verification test cases have shown that the platform complies with the requirements and that it sends the correct information to clients. Besides that, the validation test cases have shown that client applications – both web browsers and mobile devices – are able to interpret and make use of the received data. Therefore, it is assumed that this work facilitates the access to cycling information and reduces the coupling between storage and presentation.

Regarding future work, it is suggested to implement features for adding and editing data allowing cyclists to actively contribute to the growth of the database. Also, it could be interesting to study how to convert the collected cycling routes into a routable graph and how to efficiently identify, with minimum user intervention, points of interest and non-geographic information.

## Acknowledgment

The authors would like to thank CAPES Brazilian Research Agency for partially funding this work.

## References

[1] Apple, "iOS Developer Program," "https://developer.apple.com/programs/ios/ [accessed: 2012-12-13]."
[2] Biketastic, "BikeTastic," "http://biketastic.com [accessed: 2012-12-12]."
[3] CycleStreets, "Cycle routing API," "http://www.cyclestreets.net/api/ [accessed: 2012-10-22]."
[4] CycleStreets, "UK-wide cycle journey planner and photomap," "http://www.cyclestreets.net [accessed: 2012-10-22]."
[5] G. C. Parma, "Mapas cadastrais na internet: servidores de mapas," Proc. Brazilian Remote Sensing Symp. (SBSR 07), INPE, Apr. 2007, pp. 1311-1319.
[6] G. H. R. Costa, "Acesso a informações cicloviárias em dispositivos móveis: uma implementação para smartphones," Graduation Final Year Project. Dept. Comput. Sci., Santa Catarina State Univ. (UDESC), Joinville, Brazil, 2011.
[7] Gartner, "Gartner Says worldwide sales of mobile phones declined 2.3 percent in second quarter of 2012," "http://www.gartner.com/it/page.jsp?id=2120015 [accessed 2012-12-15]."
[8] GeoServer, "Features," "http://geoserver.org/display/GEOS/Features [accessed: 2012-10-22]."
[9] Google, "Download the Android SDK," "http://developer.android.com/sdk [accessed: 2012-07-25]."
[10] Google, "Google Maps/Google Earth APIs terms of service," "http://developers.google.com/maps/terms#section_10_2 [accessed: 2012-10-22]."
[11] H. D. Parker, "The unique qualities of a geographic information system: a commentary," Photogrammetric Engineering and Remote Sensing, ASPRS, Nov. 1988, vol. 54, pp. 1547-1549.
[12] J. I. Miranda and K. X. S. Souza, "Como publicar mapas na web," Proc. Brazilian Remote Sensing Symp. (SBSR 03), INPE, Apr. 2003, pp. 349-355.
[13] K. Curran; J. Crumlish; G. Fisher. "OpenStreetMap," Int. Journal of Interactive Communication Systems and Technologies (IJICST), 2012, vol. 2, no. 1, pp. 69-78.
[14] Ministry of Environment. Dept. of Climate Changes and Environmental Quality. Brazil. "1º inventário nacional de emissões atmosféricas por veículos automotores rodoviários: relatório final," "http://www.antt.gov.br/index.php/content/view/5632/1__Inventario_Nacional_de_Emissoes_Atmosfericas.html [accessed: 2012-10-22]."
[15] S. Reddy et.al., "Biketastic: sensing and mapping for better biking," Proc. Int. Conference on Human Factors in Computing Systems (CHI 10), ACM Press, Apr. 2010, pp. 1817-1820.
[16] S. T. Siddiqui. "Software Tools Required to Develop GIS Applications: An Overview," Proc. Int. Conference on Advanced Computing and Communication Technologies (ACCT 12), IEEE, Jan. 2012, pp. 51-56.
[17] S. Tu et al. "Design Strategies to Improve Performance of GIS Web Services," Proc. Int. Conference on Information Technology: Coding and Computing (ITCC 04), IEEE, Apr. 2004, vol. 2, pp. 444-448.
[18] Open Geospatial Consortium Inc. "OpenGIS web map server implementation specification," "http://www.opengeospatial.org/standards/wms [accessed: 2012-10-22]."
[19] OpenStreetMap. "Legal FAQ," "http://wiki.openstreetmap.org/wiki/Legal_FAQ [accessed: 2012-10-22]."
[20] UDESC - Santa Catarina State Univ., "Ciclovias em Joinville - SC," "http://bdes.dcc.joinville.udesc.br:100/ciclo [accessed: 2012-12-13]."
[21] World Wide Web Consortium. "SOAP version 1.2 usage scenarios: sending non-XML data," "http://www.w3.org/TR/2003/NOTE-xmlp-scenarios-20030730/#S19 [accessed: 2012-10-22]."

 ISBN: 978-1-61208-251-6 

# A New Cloud-based Deployment of Image Analysis Functionality

## The ENVI Services Engine

Thomas Bahr
Exelis Visual Information Solutions GmbH
Gilching, Germany
e-mail: thomas.bahr@exelisvis.com

Bill Okubo
Exelis Visual Information Solutions
Boulder, U.S.A.
e-mail: bill.okubo@exelisvis.com

***Abstract*—For the past 30 years Exelis Visual Information Solutions has been delivering image science tools and solutions to both research scientists and image analysts within the Environment & Natural Resources community. In this paper we present a new capability, the ENVI Services Engine, which exposes ENVI processing functionality as RESTful web services providing online and on-demand image science Apps to the geospatial community. The ENVI Services Engine is interoperable with other middleware and servers, adheres to open standards, can support multiple clients (web, mobile, desktop) and is configurable to thin-client environments, such as the JavaScript. Thus we solved the requirements of today's geospatial users expecting online, on-demand access to image processing in a client-server enterprise or cloud architecture.**

***Keywords-SDI; Geospatial Web Services; Cloud; Satellite Imagery; Image Processing; Mapping.***

## I. INTRODUCTION

Geospatial software users are looking for tools to get answers from imagery anywhere, anytime. As organizations grow and more users employ image analysis in their decision making process, these organizations have a need to centrally deploy and manage applications, tools and data. Additionally, as users are located in a variety of environments, results need to be available on thin clients such as a web browser or native mobile device [1].

Desktop software only solutions no longer meet these needs for geospatial processing. Therefore, we solved the above in an interoperable, standards-based REST (Representational State Transfer) solution.

The ENVI Services Engine (ESE) is a cloud-based deployment of image analysis functionality that integrates into geospatial platforms. The ESE is comprised of a RESTful API that allows users access to ENVI's imagery exploitation capabilities, as well as the necessary components to publish and deploy these consumable web services into any enterprise environment. By offering on-line, on-demand image analysis, geospatial users have access to information derived from remotely sensed data from any location at any time.

In the current state of development, we do not provide any off-the-shelf client application or Software development kit (SDK), leaving it free to work with any type of client.

Section II gives a short summary on ENVI, and introduces the ESE, which accesses ENVI's tools for image processing and analysis. Section III covers the development of Apps. Section IV describes a reference implementation. Finally, section V presents the conclusions.

## II. ENVI-BASED SERVICES IN THE ENTERPRISE

### A. *ENVI - Environment for Visualizing Images*

ENVI solutions combine spectral image processing and image analysis technology to get detailed information from geospatial imagery. Data gathered from today's satellite and airborne sensors, including panchromatic, multispectral, hyperspectral, radar, thermal, and LiDAR, is supported. ENVI ingests, reads, and extracts information from these various sources, and can fuse multiple data modalities to exploit the strengths of each data type. It works with any size data set and has automated tools to prepare big and small geospatial imagery for viewing or further analysis. Data preprocessing and analysis includes a comprehensive suite of tools for, e.g., atmospheric correction, registration, ortho-rectification, filtering, geospatial statistics, topographic modeling, and object based feature extraction, etc. [2].

ENVI is written in the IDL programming language. This allows users to extend or customize ENVI features and functionality to fit both image analysis requirements and specific project needs [3].

### B. *ENVI Services Engine Overview*

ESE incorporates open source standards, and, as a result, supports integration into many different types of systems. By running ENVI or IDL algorithms and routines as enterprise services, users can access any of the image analysis components they need for data exploitation. Once an ENVI or IDL routine is developed, it is quite simple to wrap it as a service and deploy it in the engine for consumption.

 ISBN: 978-1-61208-251-6 

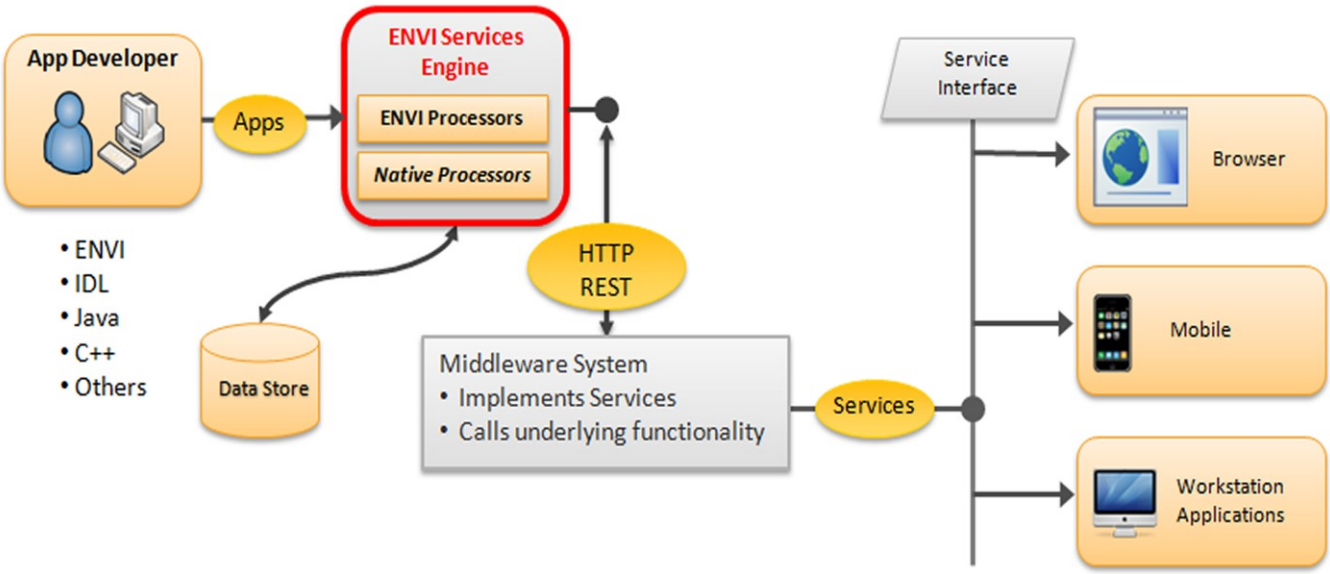



Figure 1. ENVI Services Engine conceptual rendering. ESE runs e.g. ENVI and IDL routines through a standard HTTP RESTful interface. Data requests are handled via WCS or WPS calls. ESE is called by a web client or mobile app via a middleware component.

ESE runs these algorithms through a standard HTTP RESTful interface, allowing users to provide basic user interfaces to complex analysis tasks via lightweight client. (Fig. 1). In particular, it is based on ESRI's GeoServices REST specification [4]. The ESE implements a superset of ESRI's specification wherever it makes sense for IDL and ENVI processing [5].

Results are returned to the client via standard HTTP mechanisms. They can be saved and displayed in a variety of clients, including online, desktop, and mobile clients, depending on the user's implementation. These image analysis applications and components can be integrated into a cloud environment, independent of any existing middleware configuration (Fig. 1). The ESE is therefore a flexible and easy to use framework that functions within the user's existing infrastructure.

As well as offering on-demand, online access to ENVI functionality, the ESE can be used to develop and expose unique analytics, allowing organizations to create their own enterprise image analysis workflows. Using existing ENVI functionality in a cloud environment will help to save time and resources, as the user can develop an algorithm once and deploy it to multiple locations. Additionally, the user can release new applications and functionality as mission needs change and evolve.

One of the largest benefits of deploying image analytics in the cloud is the ability to run complex, resource-intensive analysis on extremely large datasets from thin or mobile clients. By having both the data and analysis components on the web, lightweight applications can be used to call the analysis functions, making the analysis workflow more efficient for users and moving image exploitation closer to the end user.

### C. *Architecture*

The ESE application consists of one "master" process and one or more "worker" processes (Fig. 2).

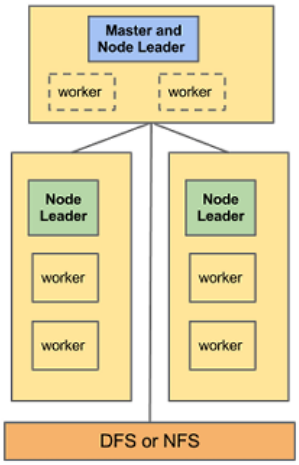



Figure 2. ENVI Services Engine architecture: "master" process and associated "worker" processes.

Processing requests are passed to the Services Engine via HTTP REST calls. The master contains an embedded web server. The master's job is to manage requests and jobs for worker processes. Worker processes contain an IDL interpreter and do the actual IDL and ENVI processing. When workers complete a job, they notify the master of the results and the master returns those results to the client [5].

ESE runs on 64-bit Linux and Windows as either a daemon or a standard application (launched manually). Security must be provided by the user by way of file and execution permissions, proxies, firewalls, etc. [5].

 ISBN: 978-1-61208-251-6

An ESE process assumes the role of master if it is the first instance of the process that is started up on the machine that is designated the master machine. On any other machine in the cluster, the first instance is run as a node leader. A node leader will launch and monitor a configurable number of worker processes. This number can be specified by a configuration file but typically defaults to the number of CPUs on the system. Node leaders, including the master, will relaunch workers that have exited [5].

### D. Scalability and Load Balancing

The important need of scalability is supported by the ability to scale and load balance because it implements a REST architecture and is designed to run on a cluster. As requests arrive from any number of clients, the master process' web server will accept and route the request. If the request is to perform a task, then the master will give the request to one of its worker processes. Worker processes can run on any machine in the ESE's cluster. The cluster is a collection of machines tied to one particular master machine. Namely, a machine is part of the cluster if it is running an instance of the ESE executable and that executable is wired to communicate with the same master as other machines on the cluster. The master machine is specified by a configuration file [5].

## III. DEVELOPING APPS

### A. Access image analysis functionality through Apps

ESE is designed for such enterprise web services and can be configured to work with multiple types of thin-and-thick clients including web browsers, mobile devices (Android, iOS) and workstation applications such as ENVI or ArcGIS (Fig. 3) [6].

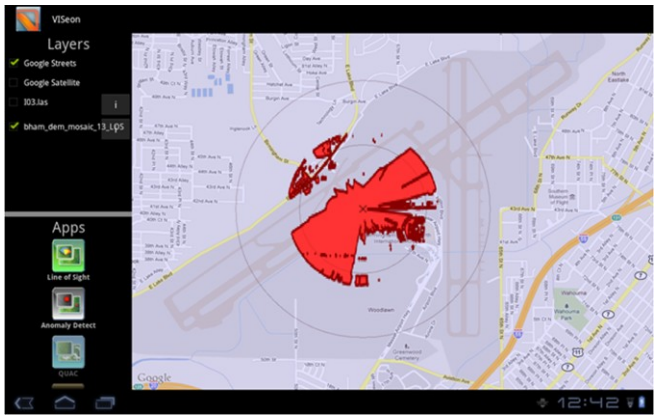



Figure 3. An interactive Line-Of-Sight App, launched from Native Android Client.
(a) The user locates an AOI, (b) draws an interactive circle to indicate parameters, (c) the middleware passes the location and AOI to the service, (d) the high-level service finds the appropriate data in the catalogue and constructs the ESE call, (e) the ESE runs and returns the vector result showing the line of sight, (f) the vector result is passed through the middleware to the client, (g) the client renders the result in the appropriate geographic location in the interface.

The advantage of this approach is that analysts can access specific image science algorithms, tools and configured workflows through discrete Apps that are easy to update and maintain.

### B. Benefits for developers creating Apps

The benefits for the developers creating Apps to use with the ESE include [6]:

- Access to 100+ different formats and modalities of imaging data.
- Ability to exploit multi-source image data using proven analytic methods such as Target Detection, Object-Based Feature Extraction, Change Detection.
- Ability to deliver derived products in easily consumable formats (.shp, kml, NITF, etc.).
- Access to analytic methods that have been widely used in the scientific community for decades.
- Ability to fuse multiple sources of information, resulting in higher confidence in derived products.
- Extensibility allows for quick insertion of capabilities to support new and emerging modalities and/or analytic approaches.

## IV. A REFERENCE IMPLEMENTATION

In an effort to provide examples of how the ESE can be used, Exelis VIS created a reference implementation that redeploys several pieces of existing ENVI code as ‘apps’. These include [1]:

- Anomaly Detection for detection of spurious material in a large image.
- Pan Sharpening for enhancing multispectral data with high resolution panchromatic imagery.
- Vegetation Delineation for identifying vegetation presence and level of vigor.
- Line of Sight to create a viewshed using terrain data (Fig. 3).
- Spectral Identification to show using a spectral library to classify materials in a hyperspectral dataset.

This implementation envisions a web client or mobile app calling the ESE via a middleware component such as MapServer, GeoServer, or some other development environment and employs a web based GUI and/or a mobile app for Android or iPhone as an access mechanism [1].

Thin and mobile clients are used to discover data using a cataloguing specification such as Web Catalog Service (WCS) and to make Open Geospatial Consortium (OGC)-compliant analysis requests via Web Processing Service (WPS) calls to the ENVI Services Engine via the middleware (Fig. 4) [1].

 ISBN: 978-1-61208-251-6 

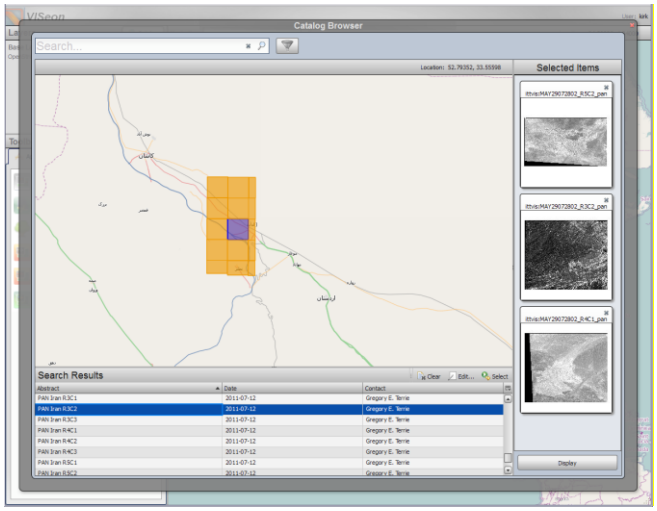



Figure 4. Accessing a data catalogue with a JavaScript interface.

One use case example could be looking at the Landsat archival data with the vegetation delineation tool. For a region or a specific scene, this ESE implementation would facilitate in discovering imagery, identifying vegetation and analyzing change over time. Results could include an animation, a profile of vegetation change over time, or specific points correlated to ground truth measurements. These results could be delivered to web, mobile or a desktop clients [1].

## V. CONCLUSION

The ENVI Services Engine (ESE) provides the community with a flexible and easy-to-use framework for creating and using image science Apps. Tasks such as detecting target signatures from hyperspectral imagery or calculating a Line of Sight (LOS) from LIDAR data are simple to execute and easy to share. The power comes from exposing all of ENVIs image processing algorithms and tools as RESTful web services and using ENVI/IDL, Java or other development languages/tools to create Apps for image science tasks such as orthorectification, classification, etc.. Benefits of the ESE approach include [6]:

- The flexibility to integrate into any architecture that can make http calls.
- Delivers an online, on-demand approach to data analytics.
- Takes advantage of open standards allowing for integration with other Apps.
- Allows experts to compile and expose their workflows and tradecraft, enabling end users to create their own tailored GEOProcessing products.
- Minimizes desktop software licensing costs.
- Emphasizes the use of server side hardware by offloading processing tasks previously performed at the desktop level.

The way imagery data is being stored and analyzed is changing. ENVI and IDL have long been tools used in Earth science communities. The ESE helps scientific data users transition to the cloud while still being able to use legacy code and algorithms [1].

The ability to host and disseminate data and complex functionality within the cloud will lead to wider enterprise-wide hosting and dissemination of earth science data in the future. The work behind ESE is aimed at making this concept a reality. More people being able to access and process scientific data means more discoveries, more collaboration, and more progress, e. g. toward future environmental challenges [1].

## ACKNOWLEDGMENT

The present contribution has been developed by Exelis Visual Information Systems in Boulder, Colorado, U.S.A.

## REFERENCES

[1] A. O'Connor, K. Lausten, B. Okubo, and T. Harris, "ENVI Services Engine: Earth and planetary image processing for the cloud", American Geophysical Union, Poster IN21C-1490, 2012.

[2] Exelis Visual Information Solutions [Ed.], "ENVI Capabilities", http://www.exelisvis.com/ProductsServices/ENVI/Capabilities.aspx, 23.01.2013.

[3] Exelis Visual Information Solutions [Ed.], "ENVI Capabilities", http://www.exelisvis.com/ProductsServices/ENVI/Capabilities/CustomizeCapabilities.aspx, 23.01.2013.

[4] ESRI [Ed.], "GeoServices REST Specification Version 1.0." An ESRI White Paper, 2010.

[5] Exelis Visual Information Solutions [Ed.], "ENVI Services Engine Documentation", ENVI Services Engine v1.0 beta, 2012.

[6] Exelis Visual Information Solutions [Ed.], "Image Science Apps for the NGA's Online GEOINT Services (OGS)." Exelis Visual Information Systems White Paper, unpublished, 2012.

 ISBN: 978-1-61208-251-6 

# An Implementation of the OGC's WFS Gazetteer Service Application Profile

## CASE: The EuroGeoNames Central Service Renewal

Lassi Lehto, Pekka Latvala and Jaakko Kähkönen
Department of Geoinformatics and Cartography
Finnish Geodetic Institute
Masala, Finland
lassi.lehto@fgi.fi, pekka.latvala@fgi.fi, jaakko.kahkonen@fgi.fi

***Abstract*—Gazetteer Service is a Web Service that provides access to geographical names data sets. The service interface has been recently specified by the Open Geospatial Consortium (OGC) as an Application Profile (AP) of the well-known Web Feature Service (WFS). The service interface was implemented in the context of the EuroGeoNames Central Service renewal project. This work represents one of the first real-life implementations of the specification and serves as a test of its adaptability in the European multi-lingual environment. The experiences are encouraging.**

***Keywords-Gazetteer Service; geographical names; geocoding; Web Services.***

## I. INTRODUCTION

### A. EuroGeoNames data set

The EuroGeoNames (EGN) data set has been originally put together in a major EU-project [1]. The main players of the project consortium included for instance the National Mapping and Cadastral Agencies (NMCA) of Germany, Austria and Slovenia and the European NMCAs' co-operation body EuroGeographics. In the initial inventory phases altogether 28 European countries took actively part in the project providing information on their place names data sets. In the end the number of countries officially committed to grant access to their national source data sets fell to a bit lower figure of 15. Still, this number of countries provides a significant share of the European geographical names content [2]. The current EGN data coverage is presented in Figure 1.

The main guiding principles of the EGN project include support for multilingual place names. A special emphasis has been put on the collection of a separate exonym database that contains names of certain important European geographical features in up to 17 different languages. The name sources of the EGN service are regarded as the most authoritative in Europe, as they are maintained by the official national agencies.

### B. EuroGeoNames service architecture

The service architecture used in the provision of the EGN content is based on a distributed model, in which each participating NMCA sets up and maintains a national EGN service instance. The centralised EGN service updates it combined geographical names data base from these local nodes. The communication between the local service nodes and the centralised service is based on the Open Geospatial Consortium's (OGC) plain Web Feature Service (WFS) access interface specification. The service architecture was initially seen as a hierarchy of cascading WFS nodes, but was eventually modified for performance reasons to include a centralised European level cache. The EGN Central Service has been operated the German NMCA, contracted by the EuroGeographics [3].

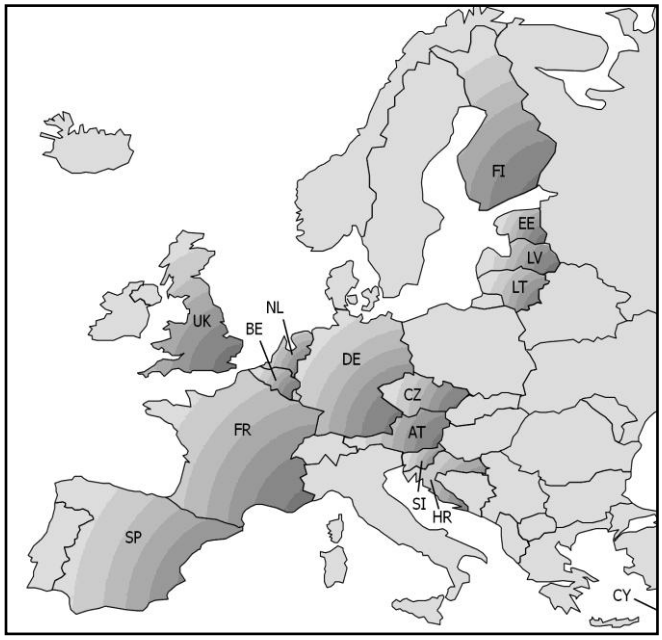



Figure 1. The current EGN coverage.

EuroGeographics has contracted the Finnish Geodetic Institute (FGI) to develop a renewed solution for EGN Central Service [7]. The main results gained so far in the still ongoing project are described in the following chapters. Motivation for the work is described in Chapter II. In Chapter III the standardization process of the Gazetteer Service interface is reviewed. Chapter IV focuses on the renewal of the EGN data model and the setup of the new Cloud Service-based EGN Central Service. Chapter V discusses in more detail the service interfaces used and explains the challenges encountered while implementing the WFS Gazetteer AP. Chapter VI contains the outlook detailing some plans for future work on the EGN data set and the service infrastructure. Finally, conclusions are drawn in Chapter VII.

 ISBN: 978-1-61208-251-6 

## II. MOTIVATION

The existing EGN Central Service implementation has been found to be outdated and in need of a renewal. The main problems with the existing Central Service are related to its highly complex database structure that is difficult to understand and hard to manage. Because of the complicated data structures, the data update procedures are difficult to execute. This has occasionally resulted in data corruption in the Central Service database, because of a fault in the update operation. In addition, the Service's output schema is considered to be too heavy for applications that operate in the Web Service environment. These difficulties have led into situation where it is difficult to convince new European countries to join the EGN service, partly because of the very demanding data structures present in the current national EGN databases. Also the use of the data set has been limited by the overly complex data structures and the too complicated access interface.

The renewed implementation of the EGN Central Service tries to alleviate the problems identified with the earlier EGN implementation. The main changes in the new Central Service are: adoption of a simplified data model and new input and output schemas, a Cloud-service based implementation architecture and the provision of new Web-friendly service interfaces.

## III. GAZETTEER SERVICE

The OGC has been working on the service interface specification for a Gazetteer service for several years. Already in 2001 a Discussion Paper was published on the topic with a title: "Gazetteer Service Draft Candidate Implementation Specification" [4]. In 2006 an official OGC Specification was published describing the Gazetteer Service as a profile of the WFS service interface, denoted by the acronym WFS-G. Already in this early version of the specification the applied information model was based on the work of the ISO TC211 on spatial referencing by geographic identifiers that has been published as an ISO international standard 19112 in 2003[5]. The latest revision of the Gazetteer Service specification has been released by the OGC as an Application Profile (AP) of the WFS in Feb 2012 [6].

According to the ISO 19112 data model a gazetteer is comprised of a set of location instances that are of a specific location type and are referenced by one or more geographical identifiers, i.e. place names. This basic principle of taking place as the basic concept and attaching potentially several individual names to a single place is in the core of the ISO Gazetteer model and subsequently one of the main challenges in the original EGN data model and database schema development.

The Gazetteer Service access interface is based on the WFS standard with a few adjustments that specify, how one can conclude that a certain WFS actually is a Gazetteer Service. The AP makes it easier to access place names data from a Gazetteer Service by explicitly defining the data model in which the place names data is to be delivered.

## IV. EUROGEONAMES DATA MODEL

The original EGN data model was developed in course of a large EU-project. Due to some design principles it grew overly complex. This was caused partly by tight adherence to a standardized way of handling multilingualism and partly because of an all-encompassing information model; a typical result of a team work in large project consortiums.

The database schema of the renewed EGN Central Service is based on the abstract ISO 19112 schema that has been fine-tuned and made concrete in the OGC's Gazetteer Service AP Best Practice document. The same ISO 19112 schema is also used as the main data input and output schema in the renewed service. The fact that the same schema is used internally in the service and also for its input and output ensures the best possible query performance and easy data maintenance.

The original EGN database structure was very complicated with many interdependencies among the database tables. As a result of the redesign and simplification the number of database tables was reduced from 22 to 6. Although a significant number of data items have been dropped, the remaining content seems to satisfy most of the typical query requirements encountered in gazetteer-related applications.

The database structure is built around the two main concepts present already in the original ISO 19112 abstract model: a location (`si_location_instance`) that can be referenced with potentially several geographical names (`alternative_geographic_identifier`). Although simple, this fundamental relationship makes the data structure to go beyond what is supported by the Simple Features data model in OGC specifications. These two object types form the core of the database, stored in two big database tables. In addition the database contains a table for storing location type information (`si_location_type`) and another table for the general information relating to the whole gazetteer data set (`si_gazetteer`). Two auxiliary tables are needed for linking the gazetteer table to location type table (`gazetteer_location_type`) and for linking the location types to an URL, from which the corresponding type classification information can be retrieved (`location_type_link`).

Multilingualism has been in the core of the EGN development since the beginning. There are three different levels in the support for different languages in the EGN service: 1. Exonyms are included into the data set for the most important European geographical features, 2. The most important attributes related to the place names are available in all major European languages, 3. The user interface of the EGN reference application is multilingual.

In the renewed EGN Central Service solution the language support has been simplified and streamlined. The exonym data set was integrated to the newly structured database unchanged. However, the original structure closely reflecting the ISO-standardized PT_Free_Text model was replaced by much simpler mechanism of maintaining language information inside the `si_location_type` table. Some multilingual data items were removed because of the

 ISBN: 978-1-61208-251-6 

simplified data model. On the service interface level the INSPIRE-standardized mechanism involving a dedicated LANGUAGE-parameter is taken into use. Location type information is encoded into the corresponding XML-element of the output schema in the requested language. In addition, the response message contains a URL-reference from which the client application can retrieve a full description of the location type, again in the originally requested language. LANGUAGE-parameter does not affect the treatment of the exonyms-related content.

User interface development for the renewed EGN Central Service is currently ongoing.

## V. SERVICE ARCHITECTURE

The renewed EGN Central Service is built according to the OGC WFS Gazetteer Service AP. It is based on the recently published OGC's Best Practice paper on the subject [2]. The national EGN Services are still hosted by the NMCAs of the EGN data provider countries for storing the geographical names data on the national level. The EGN Central Service is now maintained by the FGI as a Cloud Service instance and it contains the combined geographical names data that are collected from the national EGN Services. The Central Service is implemented as the OGC Gazetteer Service node and uses the deegree software platform as the WFS implementation and the PostgreSQL database with its spatial PostGIS extension as the database platform. The initial data upload from the national services is performed by retrieving the full datasets from the national EGN Services and replacing completely the corresponding data in the Central Service database.

Architecturally the renewed EGN implementation consists of the original national EGN service nodes, the new centralized EGN production database and the Cloud Service-based EGN Central Service that contains its own service database. Clients can connect to the service via various different access interfaces on the application layer. The EGN service architecture is thus built according to a four-layer model, see Figure 2.

The national EGN Services and the original EGN schema that they support are left untouched and their data contents are integrated to the new centralized production database that is maintained by the EGN Operator. The corresponding production database is kept as a separate data store, on which periodic updates are run by a semiautomatic process from the national EGN Services. The update process is carried out incrementally via monitoring software that queries the WFS interfaces of the national EGN Services for any updated or added data, transforms such data into the OGC Gazetteer Service AP schema and uploads the update into the production database.

The EGN Central Service is established as a Cloud Service-based OGC Gazetteer Service instance that provides access to the EGN content that is stored in the service database (EGN Cache). The service database is updated periodically by copying the production database to the Cloud. Before the update process is executed, quality checks are done to make sure the correctness and internal consistency of the production database.

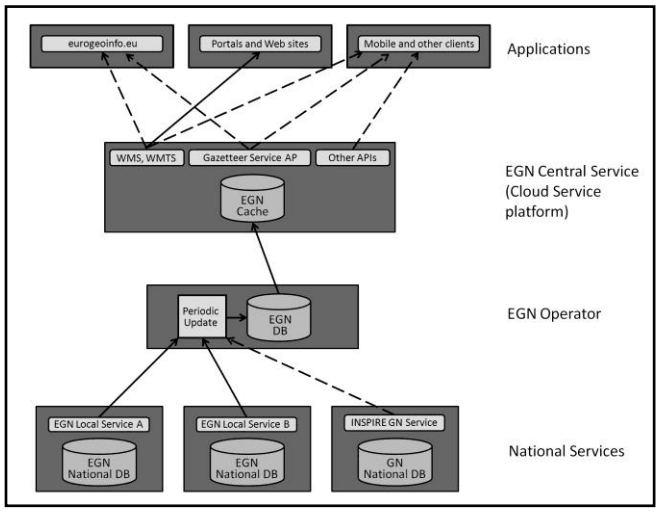



Figure 2. The new Cloud service-based EGN service architecture. The EGN Central DB maintained by the EGN Operator is periodically copied to the Cloud Service platform, from which various different client applications are supported.

## VI. FUTURE DEVELOPMENTS

In future, the EGN Central Service is foreseen to work also as an integrated source for European-wide provision of INSPIRE Geographical Names (GN) data, retrieving its data content from the national INSPIRE GN Download Services. The GN data can be updated to the EGN Central Service through a similar update procedure that is used with the local EGN nodes; by first transforming the data from the INSPIRE GN schema into the OGC Gazetteer Service AP schema and then updating the transformed data into the production database. The output will be similarly configured to support INSPIRE GN schema through an internal schema transformation process.

Initial tests on running the GN to EGN schema mapping seem promising. Most of the current simplified EGN schema can be filled in from the GN data. For update purposes it is critical that the values for the lifecycle –related attributes (beginLifespanVersion, endLifespanVersion) are available. Matching the rather coarse location type classification of the INSPIRE GN schema to the much more detailed EGN classification will be a challenging task.

The main access interface to the EGN Central Service is the interface as defined in the OGC Gazetteer Service AP. Visualisation of the EGN data can be accessed via a Web Map Service (WMS) interface. The EGN Central Service will be expanded also with other access interfaces. The probable candidates include JavaScript Object Notation (JSON)/JavaScript Application Programming Interfaces (APIs) and Keyhole Markup Language (KML) encodings.

In Figure 3 a simple client application accessing the renewed EGN service is shown. This application is connected to the service both through the Web Map Service (WMS) for the visual map display of place names, and the WFS Gazetteer Service AP interface for the results of the query.

 ISBN: 978-1-61208-251-6


Figure 3. A simple visualisation, querying and geocoding UI for accessing the EGN Central Service. The map shows EGN content displayed on top of the Google Maps satellite image backdrop. An exonym query and its results are also shown.

## VII. CONCLUSIONS

The initial experiences gained from the EGN Central Service renewal suggest that the OGC Gazetteer Service AP is good candidate for a standardized access to geographical names data content. The adopted approach of using the WFS interface as a basis for accessing geographical names content seems to be well-justified. The basic geocoding functionality is readily supported and at the same time all the available additional attribute information related to the place names can be accessed through the same interface.

The adopted ISO 19112-based information model seems to satisfy all the data requirements deemed as critical in the pre-project data simplification analysis done by the EGN Project Board. Name status attributes were relatively easy to map from the classification applied in the original EGN data to the value spaces adopted in the Gazetteer Service AP. The location type classification of the original EGN data set was used as such. A significant simplification was achieved in the database structure by reducing the number of database tables from 22 to 6.

It is seen as promising solution for the future EGN update process that the national INSPIRE GN Download Services could be used as the data source. The schema mapping from the INSPIRE GN schema was tested and it seems feasible. EGN Central Service can be also seen as a promising solution for the provision of the integrated INSPIRE GN data services with full European coverage.

Multilingualism can be supported by applying the INSPIRE-introduced LANGUAGE-parameter. This enables much more elegant and efficient way for supporting different languages than the originally applied approach of returning data in all available languages and letting the client to sort out the desired one. Additional information on the location type is available in the requested language by following a URL-link present in the response and referring back to the service.

Experiences gained in the process of deploying EGN Central Service on the cloud service platform have been entirely positive. The service has run uninterrupted for the whole test period and it is easy to make it adapt to a varying level of user demand.

The work on the EGN Central Service renewal continues. New access interfaces are being developed for the EGN Central Service. These are expected to improve the accessibility of the service and further increase its use in various application areas. The data harmonisation continues with the help of new guidelines to be developed. Coverage of the data set is expected to expand, partly helped by the new simpler data model, partly due to the use of the upcoming INSPIRE GN services.

## ACKNOWLEDGEMENT

The EuroGeoNames renewal project has been funded by the EuroGeographics Aisbl.

## REFERENCES

[1] P-G. Zaccheddu, “EuroGeoNames (EGN) – developing a European geographical names infrastructure and services,” Proc. INSPIRE Conference, Jun 23-25, 2008, Maribor, Slovenia, CD-ROM.

[2] P-G. Zaccheddu and D. Overton, “EuroGeoNames (EGN) – Implementing a sustainable European gazetteer service,” UNGEGN Working Paper No 38, 2011, at: http:// http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/26th-gegn-docs/WP/WP38_EGN_item%209_UNGEGN26.pdf [accessed: 2012-12-17]

[3] EGN Central Service, at http://www.eurogeonames.eu [accessed: 2012-12-17]

[4] R. Atkinsson [ed.], “Gazetteer Service Draft Candidate Implementation Specification”, at: http://portal.opengeospatial.org/files/?artifact_id=1041 [accessed: 2012-12-17]

[5] ISO TC211,”ISO 19112:2003 Geographic Information – Spatial referencing by geographic identifiers,” at: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=26017&commid=54904 [accessed: 2012-12-17]

[6] J. Harrison and P. Vretanos [eds.], “Gazetteer Service – Application Profile of the Web Feature Service”, OGC Best Practice, 2012, at: http://portal.opengeospatial.org/files/?artifact_id=46964 [accessed: 2012-12-17]

[7] P. Latvala, L. Lehto, J. Kähkönen, A. Jakobsson and S. Urbanas, “Renewal of EuroGeonames Central Service,” Proceeding of the INSPIRE Conference 23-27 June, 2012, Istanbul, Turkey at: http://inspire.jrc.ec.europa.eu/events/conferences/inspire_2012/index.cfm [accessed: 2012-12-17]

 ISBN: 978-1-61208-251-6 

# Design and Realization of Thematic Mapping in Web3DGIS

Guang Zhu
Beijing University of Civil Engineering and Architecture
Beijing, China
E-mail: zhuguang@bucea.edu.cn

Wenhao Qiao
Beijing University of Civil Engineering and Architecture
Key Laboratory for Urban Geomatics of National Administration of Surveying, Mapping and Geoinformation
Beijing, China
E-mail: qiaowenhao@gmail.com

***Abstract*— thematic mapping is an important module of GIS. With the development of technology such as 3D visualization and network communications, how to realize and optimize the thematic mapping based on Web3DGIS has become a hot research topic of GIS. This article has designed and realized a Web3DGIS platform based on Rich Client / Server, and adopted a thematic mapping method based on geometry. Through the method level-3 cache mechanism, the render efficiency of thematic mapping in the 3D scene and its expression ability has been effectively improved which has verified the feasibility of this method.**

***Keywords-Web3DGIS; thematic mapping; vector data rendering; level-3 cache mechanism.***

## I. INTRODUCTION

Thematic mapping, which is an important functional module of GIS (Geographic Information System), can not only render and visualize the thematic information of GIS, but also conduct further space analysis and geographical analysis. Moreover, it could exploit the hidden mode and its development tendency. With the development of technology such as 3D visualization and network communications, to efficiently realize the thematic mapping based on Web3DGIS has become a hot research topic of GIS [9]. Currently, there are mainly two kinds of thematic mapping methods with vector data and based on Web3DGIS. First, Thematic mapping method based on geometry. Second, Thematic mapping method based on texture [8].

Thematic mapping method based on geometry means rendering vector data dynamically according to the terrain surface, and the vector data exist in 3D scene as 3D feature. However, this method would cause large system overhead [3]. Thematic mapping method based on texture means transforming vector to raster data, and to load vector data into 3D scene like raster data. This method saves system overhead that theoretically can support mass data, but does not benefit for interactive data [1].

The thematic data, thematic symbol and the method of interaction are more complicated in 3D scene. ArcGIS adopts the thematic mapping method based on geometry, but it cannot support the large amount of vector data well, so the interaction capability is limited partly. Google Earth renders thematic map with KML, and rendering data with high-efficiency, however the KML limits the interaction too much. WorldWind adopts the method combining geometry and texture to render thematic map, but if the volume of data is too large, the rendering efficiency is not ideal.

In order to solve the problems which mentioned above, this article designed and realized a Web3DGIS platform based on Rich Client / Server, and adopted an optimized thematic mapping method based on geometry. Through technologies level-3 cache mechanism, the efficiency of rendering and the ability of mapping have been improved successfully in the GIS thematic mapping.

This article is organized as following: Section II presents the core idea of the solution. In Section III, the main method of thematic mapping in 3D scene via Rendering of 3D scene, Level 3 cache mechanism, and thematic data rendering are presented. Section IV focuses on the realization of the Web3DGIS, while Section V presents details on an application case for the Web3DGIS. Section VI concludes the paper.

Research purpose: 1. Enhancing thematic mapping (based on vector data) efficiency by Level 3 cache mechanism. 2. Reinforcing the interaction capability of thematic map by rendering vector data based on geometry.

## II. SOLUTION

The Figure 1 is the architecture of Web3DGIS platform.

 ISBN: 978-1-61208-251-6

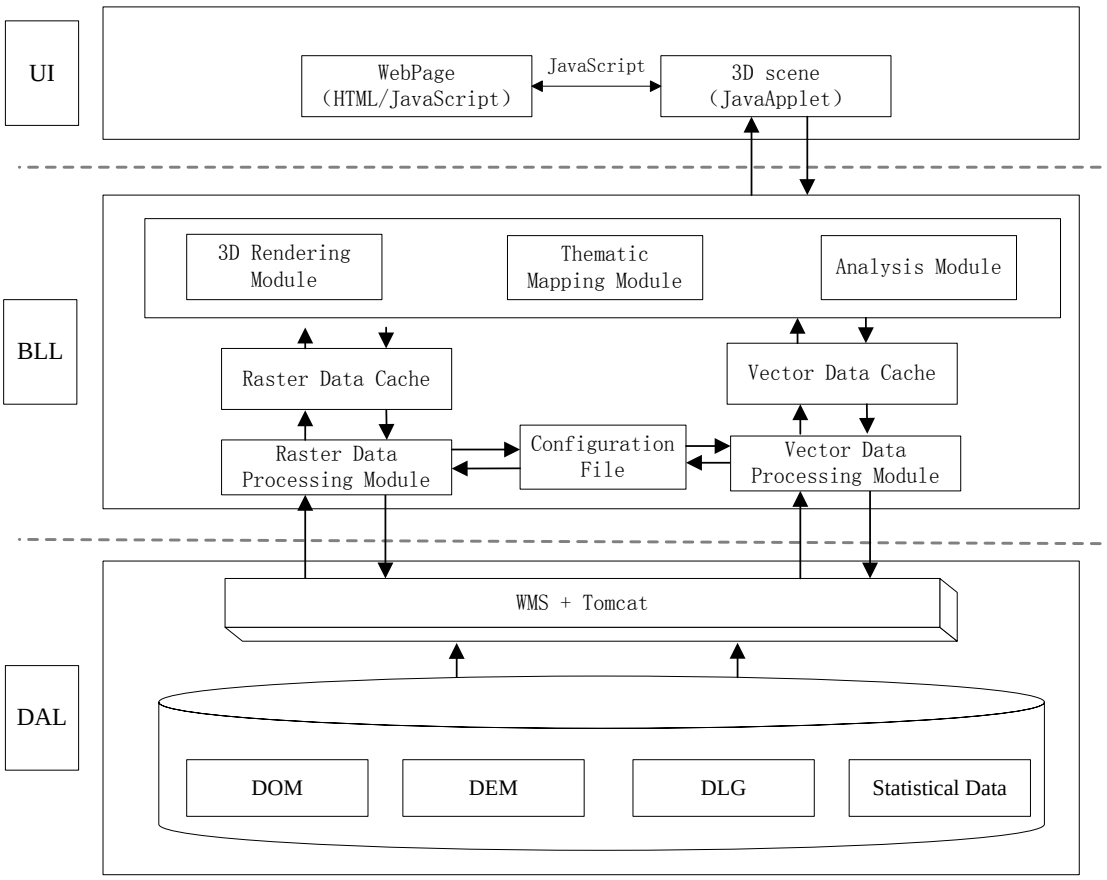



Figure 1. Architecture of the Web3DGIS platform.

The server is composed of DAL (Data Access Layer), publishes the pyramid of raster data in the database through WMS (Web map service) which follows OGC's (Open Geospatial Consortium) principles. In addition, publishes website information and vector data through Tomcat. The browser side is composed of BLL (Business Logical Layer) and UI (User Interface). After acquiring the data from the server side establishes cache specific to raster data and vector data respectively. Therein, raster data (DOM (Digital Orthophoto Map), DEM (Digital Elevation Model)) cache is used to establish 3D scene, and vector data cache is used for thematic mapping and related analysis. Considering the generality of data, the vector data in this article takes "Shapefile" as the main data source.

The BLL is the core of the architecture. After acquiring the data from DAL, which based on the configuration file, the BLL import raster data into raster data cache via raster data processing module and import vector data into vector data cache via vector data processing module. Finally, the data in the two caches will be transferred to 3D rendering module, thematic mapping module, and analysis module, in order to render and analysis of the thematic mapping.

## III. RENDERING THEMATIC MAP IN 3D SCENE

### A. *Rendering of 3D scene*

In this article, the server-side publishes LOD (Level of Detail) image tiles of DOM and DEM which adopts the WMS standard of OGC (Open Geospatial Consortium) [6]. The browser side organizes and schedules image tiles by quad tree index mechanism, loads data into the scene, establishes digital elevation model in the 3D scene by DEM tiles, renders surface images by DOM tiles, and completes the production of base map of thematic mapping and the construction of 3D scene [2].

### B. *Level 3 cache mechanism*

The thematic mapping in this article mainly depends on spatial data of vector data (.shp, .shx, .prj) and attributes data (.dbf). Through renders thematic mapping by spatial data, calculates and renders pie chart and histogram by attribute data, etc. In the mean time, this article adopts a kind of level 3 cache mechanism to make thematic mapping, in order to improve the speed of data access and rendering speed of thematic mapping. Figure 2 indicates the structure of the rendering module of thematic mapping.

In Figure 2, the vector data loader loaded vector data and transferred part of the data to vector data cache, and then the loader and the cache transferred the data to vector data

 ISBN: 978-1-61208-251-6 

complier together for compiling. At last, the compiler transferred the data stream to vector data rendering module and 3D thematic symbol rendering module to render or analyze the data.

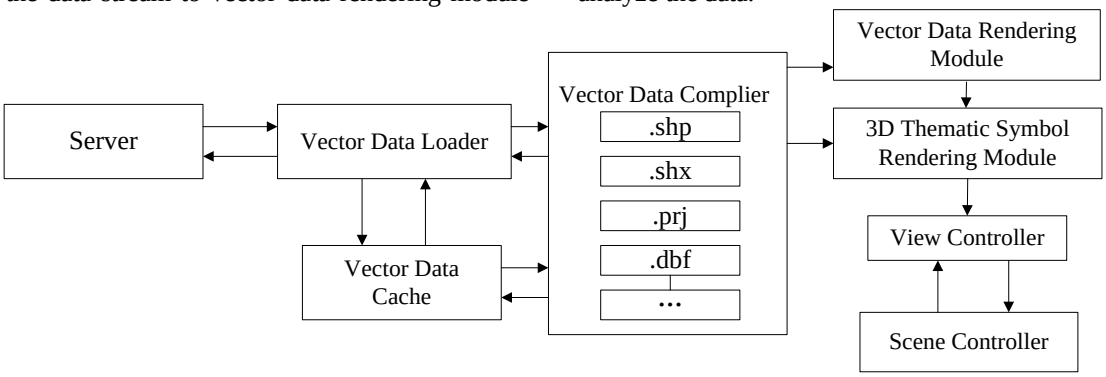



Figure 2. Architecture of the thematic mapping rendering module.

With the purpose of accelerating the speed of thematic mapping in the 3D scene, this article improves the method of active thematic mapping. Level 3 cache display mechanism is based on the traditional Level 2 cache display mechanism, for rendering thematic mapping. Level 2 cache display mechanism is to add a backup cache for organizing data to be displayed in advance on the basis of original main cache in the memory. The main cache data is used to display current scene, and the backup cache organizes data used by next scene. Level 3 cache mechanism is to add a HDD (Hard Disk Drive) cache for the interaction with server data on the basis of Level 2 memory cache. The HDD cache would be saved in the directory of browser local cache and the vector data accessed at the server side would be saved under this cache directory in the form of compressed package. Therefore, when users are to access the data again or to render again, the access request would directly go to the local data which accelerates the speed of data accession and rendering.

### *C. Thematic data rendering*

After the data being compiled by vector data compiler, based on Level 3 cache mechanism, the vector data rendering module, the 3D thematic symbol module, and scene controller will accomplish the rendering of the thematic map.

- Actually, rendering the vector data is to render the projection along the vertical direction on the terrain surface [5]. This article adopted a vector rendering method based on geometry, the core of which is rendering the vector data which is created to adapt the terrain surface. If the data adapts the terrain surface, we will compute its projection and draw it on the terrain; if not, we need to introduce some new vertexes or segments to make the data adapt the terrain surface. This way, the vector data exists in the thematic map as an "Object", while displaying the thematic map; it can also provide some analysis capability. In the meantime, there are different datasets that point to different space or time, so we should divide vector into different layers to display different thematic data by different color or symbol. At last, to reduce the loading quantity of the thematic map, controlling the visual range of the vector by the scene controller is needed. Figure 3 is the schematic diagram of rendering vector based on geometry. The broken line means the surface of terrain and the straight line means the vector data, we made the straight line adapt the broken line by introduced some new vertexes or segments.

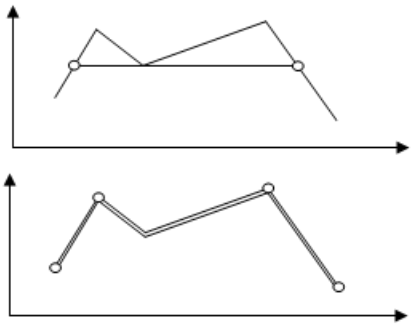


Figure 3. Schematic diagram of rendering vector based on geometry.

- Rendering thematic symbol is a kind of extension of rendering vector data in 3D scene [7]. While inheriting the method that renders vector based on geometry, we confirmed the coordinate via the spatial data of the vector data and height, color of the pie chart (histogram) via the attribute data of the vector data. In practical applications, using only figure to present the thematic map is not accurate enough; so, rendering the accurate value at the top of pie chart (histogram) is needed. The method we adopted in this article is locating the coordinate of the value to be rendered at the center of the pie chart (histogram) top surface, and overlaid the value layers on the pie chart (histogram) layers to avoid the value layers to be covered. In the meantime, the thematic symbol provides some interaction capability such as allowed user to measure or select the vector data and rendering the thematic symbol immediately, etc.

 ISBN: 978-1-61208-251-6 

## IV. REALIZATION OF THE WEB3DGIS PLATFORM

The Web3DGIS platform in this article adopted the rich client / server technology. The server is developed by C++, while integrating WMS [11] and Tomcat [12] and publishing the raster tile data and vector data, etc. The client is realized by webpage embedded a JavaApplet which is based on JOGL (Java Bindings for OpenGL). The JavaApplet takes charge of rendering the 3D scene, the thematic map and providing the application interface, and JavaScript will realize the interaction between the webpage and JavaApplet by the interface. Figure 4 is the architecture of the JavaApplet.

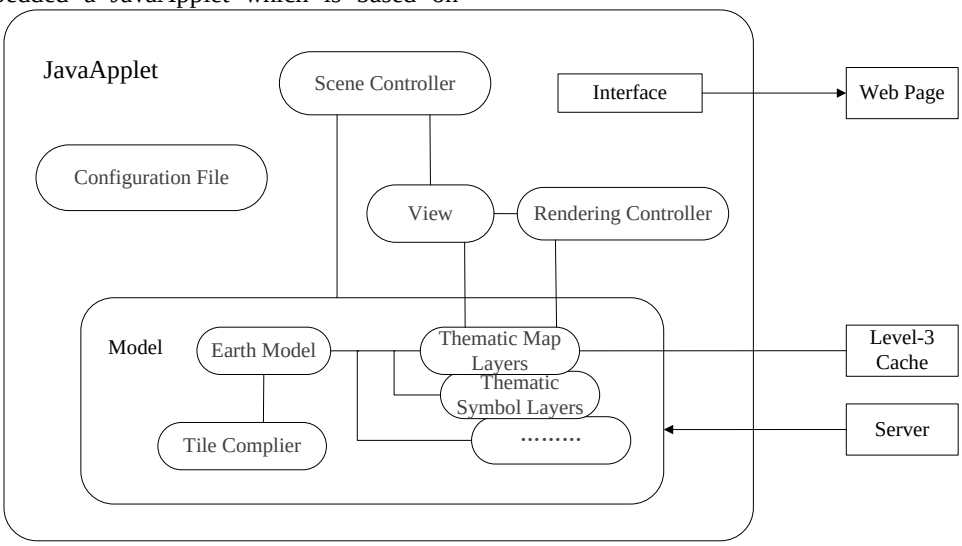



Figure 4. The Architecture of the JavaApplet.

The JavaApplet is realized by MVC (Model–View–Controller) [13][11]. The model part structured the 3D scene via raster data pyramid and loaded the thematic map in 3D scene according to spatial reference system; the view part displayed graph and transformed the coordinate from spatial to view via four-dimensional matrix; the scene controller and the render controller listened, handled the operation from user together, and displayed via the view part.

## V. APPLICATION CASE

The WEB3DGIS platform, which was designed and realized, has been used for multiple regions in different industries; for instance, a city landscape management information system. It has realized garden thematic maps in different years, different regions and different types, via rendering vector data based on geometry (Figure 5), the thematic data query via the communication between JavaScript and JavaApplet and some interactive features such as measuring the greening rate of designated area (Figure 6) via the "object" feature of vector data which is mentioned above.

Figure 5. Thematic map.

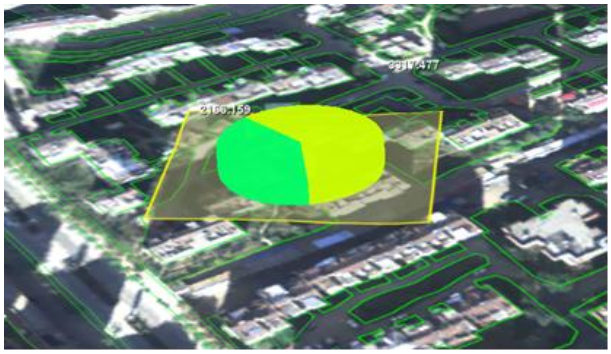


Figure 6. The design sketch of measuring the greening rate.

 ISBN: 978-1-61208-251-6

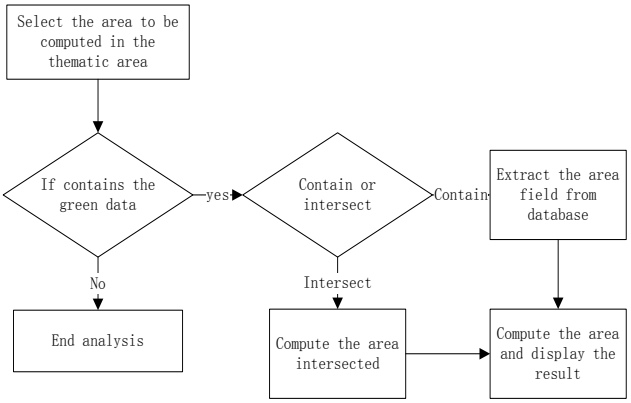



Figure 7. The flow chart of measuring the greening rate.

## VI. CONCLUSION AND FUTURE WORK

Practice has proven that adopting optimize thematic mapping and combining with the technology of rendering vector data based on geometry can realize the function of thematic mapping practically and effectively under the Web3DGIS surroundings. The frame rate can maintain in more than 100 fps while the amount of data is less than 5GB and about 60 fps while the amount of data is between 5GB to 10GB. More importantly, it can not only realize fast rendering of thematic maps, but can also realize query and analysis, based on vector data.

However, the method mentioned above cannot be well-supported for massive amounts of vector data (more than 10G). In the next step, we will embed Hilbert-R Tree into the Applet. Scheduling vector by the index or Hilbert-Value of Hilbert-R tree, in this way, the efficiency of rendering vector would be further improved.

## REFERENCES

[1] Q. Hu, "Research Based on the Integration of Two-dimensional Vector Data and Three-dimensional Terrain", Zhejinag University of Technology, 01.12.2009.

[2] H.M. Yao, T.J. Cui, S.X. Shao and J.F. Yuan, "LOD Terrain Model and Data Organization MethodBased on Quadtree Structure", Geomatics World, 2007, 12(6), pp. 56-59.

[3] P. Nazzareno, P. Fabrizio and C. Marco, "Thematic mapping at regional scale using SIASGE Radar data at X and L band and optical images", International Geoscience and Remote Sensing Symposium, IGARSS 2011 – Proceedings, IEEE, 2011, pp. 1095-1098.

[4] B. Leonhard and W. Frank, "Shallow landslide susceptibility mapping using stereo air photos and thematic maps", "Cartography and Geographic Information Science", IEEE, 2010, pp. 105-118.

[5] M. Scheider and R. Klein, "Efficient and Accurate Rendering of Vector Data on Virtual Landscapes", Journal of WSCG, 2007(1), pp. 59-65.

[6] A.S. Tang and O. Ertz, "Towards Web Services Dedicated to Thematic Mapping", OSGeo Journal, 2007(3), pp. 31-34.

[7] R. Yuri and F. Luciano, "Construction of seafloor thematic maps from multibeam acoustic backscatter angular response data"Mayer, Larry, "Computers and Geosciences", IEEE, 2012, pp. 181-187.

[8] S. Zlatanova, A.A. Rahman and M. Pilouk, "Trends in 3D GIS development", Journal of Geospatial Engineering, 2002, Vol. 4, No. 2, pp. 1-10.

[9] Y. Pan, "A Thematic Map Making Mode Based on GIS", "Bulletin of Serveting and Mapping", 2005, Vol. 9, pp. 28-40 .

[10] D. C. Roy and V. Coors, "3D web-based GIS for flood visualization and emergency response", "Society of Petroleum Engineers - 73rd European Association of Geoscientists and Engineers Conference and Exhibition 2011 - Incorporating SPE EUROPEC 2011", IEEE, 2011, pp. 1001-1005.

[11] http://www.opengeospatial.org/standards/wms: Dec. 2012.

[12] http://tomcat.apache.org: Dec. 2012.

[13] http://en.wikipedia.org/wiki/Mode-view-controller: Dec. 2012.

 ISBN: 978-1-61208-251-6 

# A Prototype for a Real-Time Indoor Evacuation Simulation System Using Indoor IR Sensor Information

Chulmin Jun, Hyunwoo Nam, Suyeong Kwak
Dept. of Geoinformatics, Univ. of Seoul
Seoul, Korea
e-mail: {cmjun, nhw612, ksykk0}@uos.ac.kr

***Abstract*—Indoor fire simulators have been used to analyze building safety in the event of emergency evacuations. These applications are primarily focused on simulating evacuation behaviors for the purpose of checking building structural problems in normal time rather than in real time situations. Therefore, they have limitations in handling real-time evacuation events for the following reasons. First, the existing models mostly experiment with artificial situations using randomly generated evacuees, while the real world requires actual data. Second, the operation takes too long to generate real time data. Third, they do not produce optimal results for use in rescue or evacuation guidance. In order to overcome these limitations, we suggest a method to build an evacuation simulation system that can be used in real-world emergency situations. The system performs numerous simulations in advance according to varying distributions of occupants. Then the resulting data are stored in a DBMS. The actual data of people captured by an IR(infrared) sensor network are compared with the simulation data in the DBMS, and the queried data most closely is provided to the user. The system has been tested in a campus building, and the suggested processes are illustrated.**

***Keywords-indoor evacuation simulation; DBMS; IR sensor; evacuation guide.***

## I. INTRODUCTION

In order to minimize damage from increasing disasters such as fire, collapse in indoor spaces, many studies on indoor evacuation models have been done to predict pedestrian's evacuation situations. Many evacuation models have been developed to model and visualize the evacuation patterns of the pedestrians. However, the primary purposes of these models are focused on building safety evaluation in such events [1, 2, 3] and they have limitations to be used for pedestrian's evacuation guide or rescue activity at the time of actual situations as follows:

First, existing evacuation models use the virtual human data to perform the simulations. However, in order to use the distribution of the real occupants, it is required for the models to have inter-communication functions with indoor sensors to find out real-time information about the present occupants in a building [4]. It is essential to have information of the present occupants acquired from the indoor sensors if the result data can be used for actual disaster situations. Second, there are time constraints for simulation results to be used in real-time situations. Since many factors, for example, physical and psychological aspects of evacuation patterns and building structures are considered in the simulation, it takes much time to generate simulation results [5]. Third, the main purposes of the most existing models studied previously have been focused on evacuation safety inspection of buildings, or have been developed for research or testing of algorithms [3]. However, if the evacuation simulation results can be used for reference data of evacuation guide or rescue operations, appropriate simulation results (*e.g.* total evacuation time, evacuation routes, the number of evacuees for each exit door, and expected bottleneck spots, etc.) should be created.

In this study, the following methodologies are proposed to solve the limitations described above. First, in order to solve the time constraints of the simulation running time, we used a method to query ready-made data instead of creating data in real-time. In our model, many simulation results are generated beforehand, and from them, results which match up with the similar conditions of the specific situations are queried and visualized. To this end, a large number of simulation results for a variety of situations (*e.g.* occupant information, simulation parameters, disaster occurrence spots, etc.) should be accumulated and stored in a system. Therefore, a DBMS-based simulation management system is required to manage a huge number of simulation results. In this study, we developed a system that can store simulation results in a DBMS. Simulation results about a variety of present occupants distribution by changing the distribution are firstly obtained, and those simulation results are stored in the DBMS. The stored data in the DBMS are transformed to be used for reference data of evacuation guide and rescue operations for evacuees. We present a process to obtain real-time placement information about occupants via inter-operation between the evacuation simulator and indoor IR(infrared) sensors, and then, to generate simulation results based on the information. The actual data of people captured by the IR sensor network are compared with the simulation data in the DBMS, and the queried data most closely is provided to the user. The system has been tested in a campus building, and the suggested processes are illustrated.

## II. RELATED WORKS

### A. Pedestrian modeling

The pedestrian behavior pattern modeling has been studied in various areas [6]. Among them, the social force

 ISBN: 978-1-61208-251-6 

model (SSM) has been used to model pedestrian movements in indoor spaces from microscopic points of views [7, 8, 9]. Also, the cellular automaton (CA) based floor field model (FFM) that generates the similar results has attracted attention [10].

In this study, the FFM-based evacuation simulator was implemented and used, which is advantageous in operation speed [11, 12]. The FFM divides indoor spaces into grid cells, and using two field values, the static floor field (SFF) and dynamic floor field (DFF), pedestrian's movements are modeled [13, 14]. In the study, the basic FFM theory and different parameter values were used and the patterns of pedestrian movements was represented by adjusting the sensitivity parameters of SFF and DFF. In particular, among the parameter values in three situations defined in [15], we used those of the cooperative regime case ($k_s$=*0.4*, $k_d$=*0.1*, $\alpha$ =*0*, $\delta$ =*0.3*) by modifying $k_s$ (0.4 to 1.0) in order to increase the moving speed of pedestrians toward exits. These parameters influence the pedestrian movements significantly, and can be set appropriately for various situations to perform simulations.

### *B. Evacuation models*

Previously, many evacuation models have been studied and developed for many purposes such as fire situation analysis, pedestrian pattern analysis and evacuation safety analysis. In [16], many previously developed evacuation simulators were compared and analyzed. In this study, data outputs generated by some representative models among them were reviewed and summarized as follows:

The analysis results showed that most of the models produced initial occupant distribution, total evacuation time, and the number of total evacuees. Also, while some models produced additional information such as building information, evacuees by hourly basis, evacuation time and route by occupants, and the number of evacuees by exit doors per hour, other models such as ASERI and EGRESS even generated the bottleneck spots. Egress Section in FPETool, Simulex, and EXIT89, which were applicable to be used in multi-layered space structures, produced information such as stair information and vertical travel time. The evacuation simulation system proposed in this study should produce the resulting data which is compact to be stored in the DBMS and highly applicable to the actual disaster situations.

Also, existing developed evacuation simulators such as Simulex and buildingEXODUS were mostly commercial programs, which were impossible to modify or add new features. In this study, since generation of the required resulting data and storage of the simulation results in the DBMS were needed, which were not present in the existing evacuation models, an evacuation simulator including the required features should be developed based on the one developed by the authors [17].

### *C. Indoor occupants detection sensor*

In order to find out the indoor occupants distribution or the number of indoor occupants, a variety of sensors such as Infrared (IR) sensors, RFID, heat detection sensors, weight detection sensors, Zigbee sensors, and camera sensors have been used. In [18], using an indoor camera sensor, a method of acquiring the location information of moving objects (pedestrians) was proposed. In [19], using the video information acquired by CCD cameras installed at the exit doors, information of moving objects was obtained too. In [4], a detection system was proposed to find the number of pedestrians and moving direction by measuring heat generated by pedestrians using the IR distance measurement sensor. In [20], a system to trace moving objects based on IR was proposed.

The required information in this study was the distribution of occupants by each room, and accurate location information for each occupant was not needed. Since distribution information of the occupants based on the number of the occupants in each room or hallway was needed, exit or entry information of the occupants was only acquired. A sensor which is appropriate for acquiring the exit or entry information of the occupants is an IR sensor. In this study, a sensor network was configured to collect the actual occupants data by installing IR sensors at certain areas of the campus buildings, which will be described in detail in the next section.

## III. SYSTEM DEVELOPMENT

### *A. Generation and storage of the simulation results*

In this study, a large number of simulation results were stored in the DBMS beforehand, and a system was implemented to search and output resulting data for a specific situation through queries. For this, resulting data which can be stored in the DBMS should be selected first, and then a database schema for them should be designed. As discussed in Section II, evacuation simulations were performed using the FFM with some modified parameters in this study [17]. The resulting data to be generated and stored were summarized in Table I.

For more explanation of Table I, basic storage information included initial occupants' distribution, total evacuated occupants and time, and algorithm parameters. For

TABLE I. TYPES OF THE GENERATED DATA

| Resulting data | Information to be stored in DBMS |
|---|---|
| Initial occupants distribution | Initial location coordinates for each pedestrian |
| Total evacuated occupants | Total evacuated occupants |
| Total evacuation time | Total evacuation time |
| Algorithm parameters | The algorithm parameters used in the FFM |
| Evacuated occupants by exit door | Evacuated occupants collected by each exit door |
| Evacuation route by each room | Coordinates list for the travelled route by selecting a representative occupant for each room randomly |
| Expected bottleneck spots | Cell occupancy degree value for each cell by calculating pedestrian's occupancy |

 ISBN: 978-1-61208-251-6 

the initial occupants distribution, simply put, the number of occupants in each room was stored, and initial position for each pedestrian can be stored as cell coordinates in more detail. The total evacuated occupants and time, which are the information provided by most evacuation simulators, were suitable to be stored in the DBMS. The algorithm parameters mean the parameter values of $k_s$, $k_d$, $\alpha$, and $\delta$, which were applied to SFF and DFF in the FFM. These values should be stored because evacuation patterns changed greatly by changes of these values [15].

In this study, together with basic information, additional information which can be required for actual disaster situations was produced as the resulting data which was then stored in the DBMS. The evacuated occupants by exit doors were collected and stored by calculating evacuated occupants by each exit door during the evacuation simulation running. The evacuation route for each room was stored by selecting a representative occupant of each room and storing his/her travel coordinates during evacuation. The reason for the selection of the representative for each room was to minimize data size otherwise too large to be stored in the DBMS. Finally the expected bottleneck spots represented information about bottleneck which can be occurred during evacuation at the hallways, doors or exits. In this study, the number of pedestrians occupied in each cell was recorded and this data was called the cell occupancy degree. If the distribution of the cell occupancy degree in an indoor space is known, then areas where pedestrians were crowded can be estimated. Figure 1 shows the visualization of the cell occupancy degree. The figure represented the cell occupancy degree by dark colors and showed that occupancy was most frequent at the lower exit. Based on this information, it can be estimated that which areas are likely to have bottleneck.

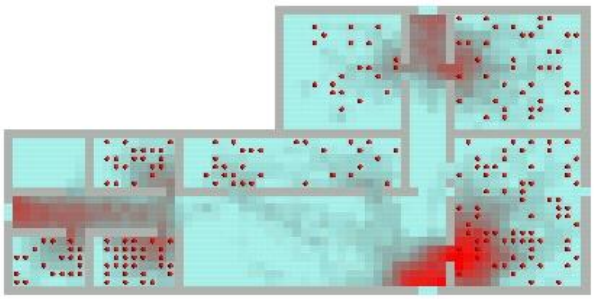


Figure 1. Example of cell occupancy

TABLE II. DBMS SCHEMA

| Table | Schema |
|---|---|
| Sensor Detection Occupants | ID, Person, Type(FK), Time, TotalPerson |
| Sensor Type | ID, TypeDescription |
| Simulation Results | ID, Building, Person, TotalPerson, Exit, Time, Alpha, Delta, Kd, Ks, SecPerTick, AgentPathLog, CellOccupationValue, PersonLocation |

In order to store the resulting data described above in the DBMS, a schema should be designed. Table II shows the table schemas in the DBMS. It consisted of three tables of sensor detection occupants, sensor type and simulation results. The sensor detection occupants table can store the number of detected occupants by hour and the recorded time as well as the total occupants. The sensor type table is information about the sensor location in which space it was installed. The simulation results table was designed as described above.

### B. *Capturing indoor occupants distribution using the exit and entry detection*

In order to utilize the simulation data in an actual disaster situation, simulation results generated using the actual occupants distribution are needed. Therefore, sensor technologies applicable in indoor spaces should be used to find out the indoor occupants distribution. As an appropriate sensor for capturing indoor occupants distribution, infrared sensors were selected. The data which were transferred to the server through the sensor were detection time, detected exit and entry occupants in real-time, and the number of cumulative occupants.

The transferred data were designed to be stored in the DBMS. Since a connection to the sensor can be disconnected when an actual disaster occurs, the system was designed that the exit or entry occupant data detected by the sensors were stored in the DBMS in the server first, and then the occupants distribution information at the specific time was queried and fetched.

### C. *Matching between the evacuation simulation results and the detected occupants*

Section 3-A showed how the resulting data to be generated for the evacuation simulation were defined, and a database schema for the simulation results was designed to store the data. Section 3-B summarized that an infrared sensor network was built and detected occupants information was stored in the DBMS.

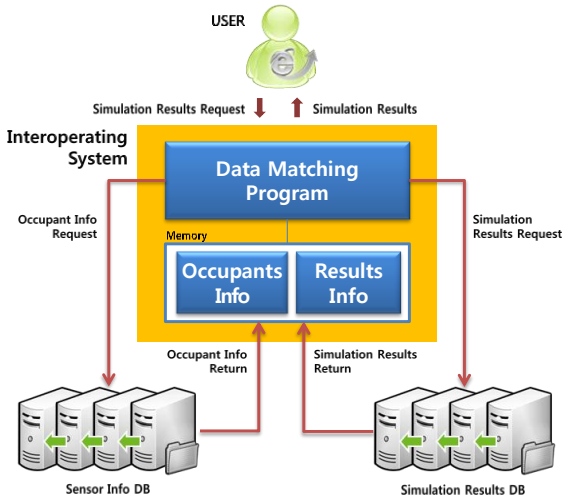
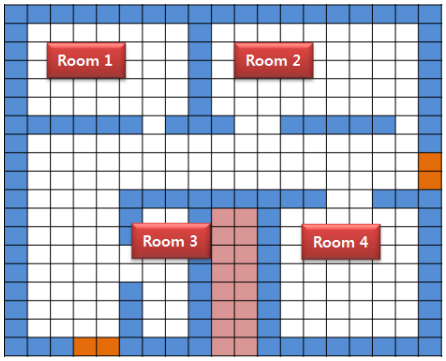



Figure 2. Configuration and flow of the matching system

 ISBN: 978-1-61208-251-6 

Figure 2 shows a configuration and data flow of the matching system between real-time occupants data by sensor detection and the simulation results proposed in this study. The interoperating system has a data matching program, which performs queries and returns the query results by connecting to the simulation results database that stores the simulation results and the sensor detection occupants database that stores the exit and entry occupants information detected by the sensor. The sensor detection occupants DB stored the exit and entry occupants information of each room detected in every 5 minutes. The simulation results DB stored the simulation results generated beforehand according to various combinations of occupants. In this study, among the various factors which influence the simulation results, the occupants distribution was focused so that simulations were run by changing combinations of occupants and its results were stored in the DB. Figure 3 shows an example of different combinations of occupants distribution. After all the possible combinations of distributions, each of which is composed of a set of occupants numbers in building rooms, were established, a simulation was run for each case. Then, the results were stored in the database table.

To explain the system's flow, first, a user requested a simulation results at the specific time such as the disaster occurrence situation. Then, the interoperating system requested the occupant information from the sensor detection occupants DB in order to obtain the occupants combination information at the closest time with the requested time. The requested occupant information was returned to the system, and stored in the memory. Next, in order to have the simulation results which were performed with the closest occupants combinations of the obtained occupants distribution information, the results were requested from the simulation results DB. Then, requested simulation results were returned to the system, and finally delivered to the user.




| Case | Room 1 | Room 2 | Room 3 | Room 4 |
|---|---|---|---|---|
| Case 1 | 5 | 0 | 0 | 0 |
| Case 2 | 5 | 5 | 0 | 0 |
| Case 3 | 5 | 5 | 5 | 0 |
| Case 4 | 10 | 5 | 5 | 5 |
| Case 5 | 10 | 0 | 5 | 0 |
| Case 6 | 10 | 10 | 5 | 10 |
| Case 7 | 0 | 0 | 10 | 5 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

Figure 3. Different combinations of occupants distribution

## IV. SYSTEM TEST

### A. *Development of the DBMS and the system*

A DBMS was used to store the simulation results and the sensor detection occupants information as described in the previous section. In this study, PostgreSQL was used as the DBMS [21]. The DBMS consisted of three tables such as IRSensor(storing occupants information detected by the sensor), SimResult(storing the simulation results) and SensorType(storing the configuration type of the sensor). Each schema for the tables was configured as described in Section III.

Next, based on the evacuation simulator which was developed by [17, 22], the revised evacuation simulator proposed in this study was developed. There were two features required for the evacuation simulator: one is an automatic iteration by changing various occupants combinations, and the other is to store the result data in the simulation results DB after completion of simulation running.

Using the developed simulator, simulation results according to the various occupants combinations were stored in the DBMS. Since it was not practical to constitute all the possible combinations from occupants of rooms, some increments (i.e. 5 as shown in Fig. 3) were used. Then, using the developed DBMS, the matching system was developed to fetch the simulation results. This system followed the configuration and flow as explained in Section III-C. To obtain the sensor detection occupants information at the specific time, queries with time information were used. The numbers of occupants stored in the DB, with which simulations were run, were integers with some increments as shown in Figure 3 (*e.g.* 5, 10, 0, 5, etc.), while those captured by the sensors were the actual numbers of people located in rooms (*e.g.* 4, 7, 1, 2, etc.). Therefore, it was not possible to expect to find two sets that match exactly. To find the simulation results generated by using the occupants distribution similar with the captured occupants information by sensors, the Euclidean distance was used, which calculates the difference between the number of the occupants of each room and the one that was captured by sensors as

$$\mathrm{Dist}(d_i, d_j) = \sqrt{\sum_{k=1}^{n} \left(w_{ik} - w_{jk}\right)^2} \quad (1)$$

where $d_i$, $d_j$ are the occupants distribution in the DB and the sensor, and $W_{ik}$, $W_{jk}$ are the number of occupants in room $k$ in the DB and the sensor respectively. The lower the value, the higher the similarity between the simulation results in the DB and the occupants information from the sensor at the specific time. Then top 20 result data were displayed to the user. Basically, this information was provided in a text format, and additional information was displayed through the detailed information window.

 ISBN: 978-1-61208-251-6 

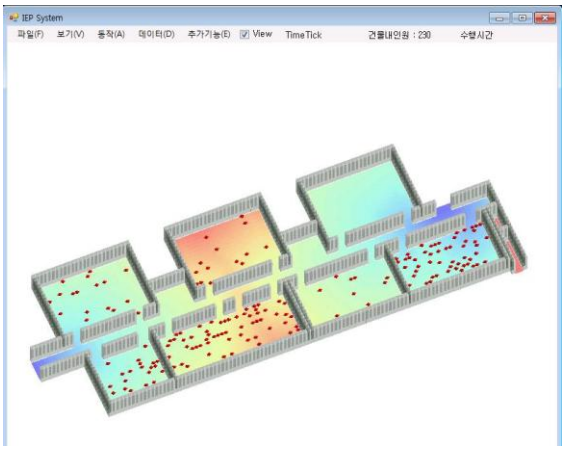


Figure 4. Main interface of the evacuation simulator

### B. System test

To evaluate the developed system, a part of the campus building was selected as the test area. It consisted of seven lecture rooms and a hallway, and infrared sensors were installed at all doors to detect the exit and entry occupants in the lecture rooms.

Figure 4 shows the main interface of the evacuation simulator used in the study. The figure displays the interface before input of building data and simulation running. When simulation was running, virtual evacuation was performed and the resulting data were stored in the DBMS. Approximately, 20,000 times of simulations were performed. After simulations and database storage were finished, the matching process was performed. Figure 5 shows the main interface of the matching system. The figure displays the result after matching was done with the occupants data detected by the sensors. According to the calculated scores, top 20 results were generated with basic information as a text format. The time taken for search execution was approximately 0.5 second, which showed very short time compared to the actual simulation running time.

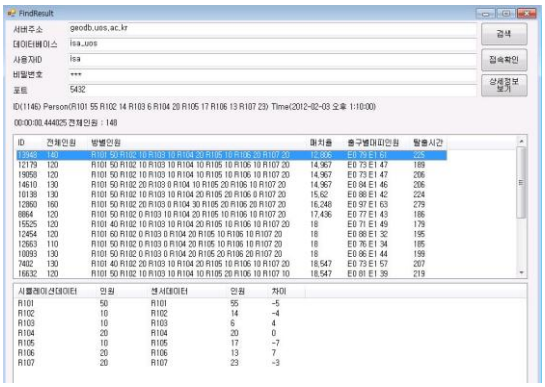


Figure 5. Main interface of the matching system

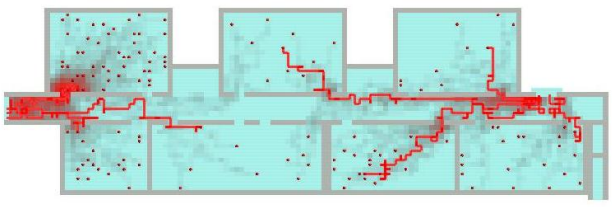


Figure 6. Visualization of the result data

Figure 6 shows the visualized scene of the result data. The points represent the initial positions of the occupants, and the dark colors in the floor represent expected bottleneck spots as indicated by the cell occupancy rate. The figure showed more occupants were crowded in the left hallway. The lines represent the evacuation routes which were traveled by an occupant of each room which was randomly chosen. By these lines, it can be estimated that which routes were chosen to evacuate for occupants in each room.

## V. CONCLUSIONS

In this study, a real-time evacuation simulation system was developed by using indoor sensors and DBMS. By using infrared sensors, the actual distribution information of occupants was detected and applied. Also, the evacuation simulation results were stored in the DBMS, and the simulation result data, which was corresponding to the occupants distribution at the specific time, was provided by matching with the sensor detection occupants. In addition, types and configuration for the simulation result data were defined, and a visualization module for the result data was developed.

In this study, we demonstrated that evacuation simulation results can be used in not only evacuation safety evaluation of the building but also application to the actual disaster situation through our proposed system. Using the indoor sensor technology and obtaining simulation results through fast DBMS matching, limitations of the existing models inapplicable to real-time events were overcome.

In this study, a sensor network was configured, and evacuation simulation results were obtained by focusing on the occupants distribution combination. In the future, not only using the occupants distribution combination but also by capturing the characteristics of the occupants (*e.g.* age, sex, disposition, etc.) via sensors, our proposed system can be improved further to obtain the simulation results which incorporate the characteristics and distribution of the occupants.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF- 2009-413-D00001).

## REFERENCES

[1] So-Yeon Kang, Jeong-Hoon Cho, Sun-Ho Choi, Min-Ho Hong, and Won-Bok Jang, "Assessment of evacuation safety in hospital by use of simulation", The Society Of Air-Conditioning and Refrigerating Engineers of Korea 2004 Summer Conference, Korea, 2004.

 ISBN: 978-1-61208-251-6 

[2] Jong Hoon Kim, Woon Hyung Kim, Jun Ho Hur, and Sam Kew Roh, "The evaluation of evacuation performance on large-volume spaces", Korean Institute of Fire Science & Engineering 2003 Autumn Conference, Korea, 2003.

[3] Ho Ju Yoon and Yen Kyung Hwang, "A study on the research trends analysis for the eevelopment of evacuation simulation program", Korean Institute of Fire Science & Engineering 2009 Spring Conference, Korea, 2009.

[4] Kee-Hwan Ka, Gwang-Gook Lee, Ja-Young Yoon, Jae-jun Kim, and Whoi-Yul Kim, "People counting using infrared distance measurement sensor", The Institute of Electronics Engineers of Korea 2008 Summer Conference, Korea, 2008.

[5] Ho Ju Yoon and Yen Kyung Hwang, "Studies in comparative analysis on algorithm of Exodus and Simulex for the evacuation simulation program development", Korean Institute of Fire Science & Engineering 2009 Spring Conference, Korea, 2009.

[6] R. K. Ahuja, T. L. Magnate, and J. B. Orlin, Network Flows: Theory, Algorithms and Applications, 1993.

[7] D. Helbing, I. Farkas, P. Molnár, and T. Vicsek, "Simulation of pedestrian crowds in normal and evacuation situations" In M. Schreckenberg and S. Sharma (Eds.), Pedestrian and Evacuation Dynamics, Springer-Verlag, Berlin, pp. 21-58, 2001.

[8] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic", Nature, vol. 407, pp. 487-490, 2000.

[9] D. Helbing and P. Molnár, "Self-organization phenomena in pedestrian crowds", In F. Schweitzer (Eds.), Self-Organization of Complex Structures: From Individual to Collective Dynamics, Gordon & Beach, London, UK, 1997.

[10] A. Kirchner and A. Schadschneider, "Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics", Physica A, vol. 312, pp. 260-276, 2002.

[11] C. Henein and T. White, "Agent-based modeling of forces in crowds", Springer, vol. 3415, pp. 173–184, 2005.

[12] C. Henein and T. White, "Macroscopic effects of microscopic forces between agents in crowd models", Physica A, vol. 373, pp. 694-712, 2007.

[13] K. Nishinari, A. Kirchner, A. Namazi, and A. Schadschneider, "Simulations of evacuation by an extended floor field CA model", Traffic and Granular Flow '03, pp. 405-410, 2005.

[14] T. Kretz and M. Schrekenberg, "Floor field- and agent-based simulation tool", International Symposium of Transport Simulation, Lausanne, Switzerland, 2006.

[15] A. Kirchner, K. Nishinari, and A. Schadschneider, "Friction effects and clogging in a cellular automaton model for pedestrian dynamics", Phys. Rev., E 65(5): 056122-056128, 2003.

[16] E. D. Kuligowski and R. D. Peacock, "Review of building evacuation models", Report NIST TN 1471, Gaithersburg, MD: National Institute of Standards and Technology, 2005.

[17] Suyeong Kwak, Hyunwoo Nam, and Chulmin Jun, "An indoor pedestrian simulation model incorporating the visibility", The Journal of Geographic Information System Association of Korea, Vol. 18, No. 5, pp. 133-142, 2010.

[18] Hyun-jin Lee, So-young Park, Soon-ju Yoo, Ea-hwa Kim, and Chulmin Jun, "Real time mapping system for moving objects based on indoor camera sensor", Geographic Information System Association of Korea 2008 Spring Conference, Korea, 2006.

[19] Chang-yul Kim and Seok-rim Choi, "A camera-based system for counting people in real time", The Institute of Information Technology Assessment 2002 Autumn Conference, Korea, 2002.

[20] Seokmin Jung, Woojin Jung, and Woontack Woo, "Infrared - based user location tracking system for indoor environments", Journal of the Institute of Electronic Engineers of Korea CI, Vol. 42, No. 5, pp. 9-20, 2005.

[21] PostgreSQL, http://www.postgresql.org/.

[22] Inhye Park, Chulmin Jun, and Jiyeong Lee, "GIS-based fire evacuation simulation using CA model", The Journal of GIS Association of Korea, Vol. 16, No. 2, pp. 157-171, 2008.

 ISBN: 978-1-61208-251-6 

# Guiding the Controller in Geovisual Analytics to Improve Maritime Surveillance

Gabriel Vatin, Aldo Napoli
MINES ParisTech
Centre for research on Risk and Crisis (CRC)
Sophia-Antipolis, France
gabriel.vatin@mines-paristech.fr, aldo.napoli@mines-paristech.fr

***Abstract*—Maritime traffic surveillance requires a very accurate and continuous analysis of the sea. This area consists of many different objects, actors and rules. Monitoring such a wide and complex area requires adapted visual tools, such as Maritime Surveillance Systems. These tools help identifying abnormal behaviours, which can lead to risky situations. We investigate the usability of visual analytics and geovisualization methods that will improve these systems: a better synthesis of the data and more effective tools lead to improved situation awareness. Visualization methods and needs of controllers being very specific, there is no single solution for modelling, visualizing and analysing maritime data. In this paper, we identify the limits of current research in geovisual analytics for maritime surveillance. A new approach for guiding the selection of visual analytics methods is proposed. The profile of the user, the purpose of use and the situation to analysis are considered together; a knowledge-based system will guide the user toward the most suitable visualization methods.**

***Keywords-Geovisual analytics; geovisualization; maritime surveillance; use and user issues.***

## I. INTRODUCTION

Maritime traffic surveillance asks for many stages of control. This includes monitoring the traffic, detecting anomalies like abnormal behaviours, and finally decision making to preserve security (security of persons, pollution prevention) and safety (illegal acts prevention and fight) at sea and along the coast. Having an accurate knowledge of what is happening in the area of interest (current situation) and understanding how events will evolve in the near future (projection) are key principles to situation awareness [1].

Surveillance requires many heterogeneous and dynamic data, in order to control maritime situation: Automatic Identification System (AIS) coupled to marine radar provides real-time information about ships position, speed and course by radio. In order to display these data, controllers and analysts use Maritime Surveillance Systems (MSS). Dynamic data are merged with static data such as nautical charts and meteorology layers.

Figure 1 gives an overview of vessel traffic data that have to be monitored in real-time. On this image, triangles stand for moving ships, whereas squares are static ships. The colour represents the type of boat, broadcasted by AIS data (e.g., *Fishing*, *Tanker*, *High Speed*, etc.). In this paper, the images illustrating cartographic representations of vessels are captured from our web-mapping system called FishEye. It is used to test visualization with web libraries and OpenLayers mapping.

But the poor display of MSS does not suit the human needs for situation awareness. On the one hand, the amount of data, human stress and cognitive limits of vision make decision process even more difficult [2], [3]. On the other hand, MSS do not offer a real analysis of the data with effective tools: vessels trajectories and positions are only displayed, without complete analysis tools. Yet, the users need the most efficient tools to control maritime system, without an entirely automated process [4], [5].

Therefore, visual methods have to be used to let users analyse the data and extract knowledge. In this paper, we propose an approach for guiding the user in selecting the most appropriate visualization environments, according to his / her profile and the tasks to perform. The limits in traffic visualization are pointed out from literature, and are used as a basis for our own research. Therefore, the main stages of our approach are explained, for improving visual data exploration in maritime surveillance systems.

Perspectives in developing geovisual analytics taxonomy are presented in the last section: it will be used for formalizing expert's knowledge in geovisualization.

This paper is structured as follows: Section II presents maritime surveillance systems and related work on traffic anomaly detection, particularly maritime traffic. Section III describes the needs in a general control system that uses cognition in a human / machine environment, and the way geovisual analytics fulfil these needs. In Section IV the basis of a research methodology is proposed for solving the problem. Finally, Section V concludes the paper.

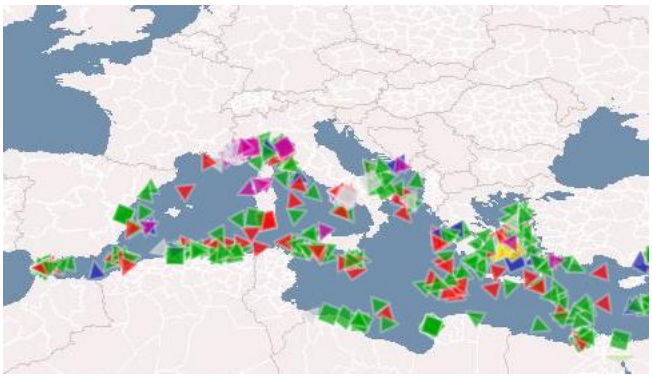


Figure 1. Overview of real-time traffic data in the Mediterrannean Se, with AIS data (FishEye, MINES ParisTech - CRC).

 ISBN: 978-1-61208-251-6 

## II. RELATED WORK

### A. Analysis of Maritime Data

Maritime Surveillance Systems (MSS) are cartographic platforms that display the position of vessels on top of electronic navigational charts. These tools allow operators and analysts to monitor a zone of interest and to store the history of vessels movements. But due to the large amount of displayed data, detecting suspicious or dangerous behaviour happens to be very difficult for human operators; therefore analysing maritime data is a real challenge for human beings.

The literature provides several methods to display and analyse maritime information, and to prevent the risks by automatic or semi-automatic means, which would be one of the three types: modelling the data, processing the data or investigating the interface. An example of modelling is using spatial ontologies based on the formalization of the knowledge of experts to detect anomalies in vessels movement, within real-time data [6]. Trajectories and events modelling must be chosen before being analysed (e.g., [7], [8]).

Spatial data-mining is used beforehand in order to extract rules and scenarii from past events, for the processing part [9]: the knowledge of experts is used to complete these results. In the process of decision making, Bayesian networks help users to take relevant countermeasures [10], based on data-mining and brainstorming results. Multi-agent systems have also been investigated in [11] to simulate ships trajectories and their behaviour while coming across another vessel.

To include human actors in this process and improve situation awareness, much research was led on the human / system interface. The use of geovisualization methods is a major issue for traffic surveillance: mapping historical data provides visualization of various behaviours at sea [12] and using statistical methods to compare actual data to a normal model allows highlighting anomalies [13].

We chose to improve the previous work that was done about geovisualization and visual analytics for supporting risk detection and risk management. Indeed, the literature and the projects we led previously showed that human role is still neglected in risk management, for the benefit of artificial intelligence [13], [14].

### B. Visual Analytics for Traffic Control

We have seen that the interfaces of MSS do not propose complete interactive and visual methods for accurate data analysis: their design is suitable for a single user looking for a global view of maritime domain and querying some information on ships. But there is a gap between a system for a single user and a system for decision making, designed for various users and various needs (close to the geocollaboration field). According to MacEachren and Brewer, visualization systems and GIS were first developed for a single user [15], [16]: MSS present the same limit, since they are based on a GIS general framework. Therefore, systems that have to be used for crisis management, which involve various user profiles and many different analyses to be led, ask for specific collaboration and information exchange tools [17], [18], [19].

The first user profile to be considered is the person who is directly in the studied environment: sailors, captains, pilots, etc. As they are at the heart of the controlled system (e.g., the sea, the sky), they do not have the same perception of the environment than an external controller would have. In order to take this point of view into account, studies have been done in modelling and displaying of trajectories with user-centred view [7] and augmented reality [20]. The results of these studies are methods and interfaces to help actors to analyse real-time data with their own point of view, which lead to a better situation awareness.

Other actors, like traffic controllers or analysts, work with real-time or past data: they need a more global view on the system to extract patterns and monitor disturbances [21], [22]. Geovisual analytics tools provide means to explore large sets of data and to analyse both *overview* and *detail* scales using maps and graphs [23], [24]. By combining various methods and tools, as [25] did with *Triple Perspective Visual Trajectory Analytics* (TripVista), traffic data can be analysed at different scales in order to extract patterns or abnormal behaviours.

Reference [26] presents a state of the art of existing software and methods for information visualization and its application to maritime surveillance. It gives an overall view of all possible needs for this domain and the type of methods that fulfil them: visualizing space-time data, discovering unknown information, displaying uncertainty of data, etc. However, these visual analytics environments may require skilled users to manipulate it and really extract information from the data. The authors conclude the report by reminding visualization problem in maritime security that should be automatically proposed: *Visualizing coverage and ignorance*, *Visualizing ship tracks in time and space*, *Visualizing "normal" behaviour of ships* and *Visualizing attribute data of interest*.

Respecting these points, new means of visualizing information and analysing past or real-time data have been developed recently in maritime and aerial domain. Willems used composite density maps to highlight various behaviours [12], whereas Riveiro led her research about detecting unusual behaviours at sea with self-organizing map and Gaussian mixture models [13].

Latest research highlighted the role of geovisual analytics in traffic surveillance. Methods and tools have been developed to support anomalies detection and situation awareness. These studies provide major results for improving control systems using geovisualization. However, the aims and the functionalities of control systems in general need to be further investigated: this would lead our forthcoming research in the role of geovisual analytics for risks management.

In the next section, we introduce the definitions of risk and control, and we discuss the limits of previous work according to this concept.

 ISBN: 978-1-61208-251-6

## III. COGNITIVE SYSTEMS FOR MARITIME CONTROL

In the field of risk management, whether it is in a company or on a geographic territory, four main steps must be controlled [27], as shown in Figure 2. Potential risks must be known and anticipated to prevent them. *Vigilance* stage is about controlling the system and monitoring what is happening. Then, if unexpected events happen, a solution must be found to stabilize the system (e.g., search and rescue mission). Finally, this cycle is improved with the analysis of past events (feedbacks), which feeds the anticipation step with new rules.


Figure 2. Risk management process according to Wybo. This cycle is the basis for control systems.

Within maritime surveillance, the operators are in charge of monitoring the traffic, which is mostly the *Vigilance* step. With the help of analysts and other actors, the three other stages are led with group decisions and analysis. The stage called *Handling with unexpected* requires an accurate analysis of the situation, so that actors could take a suitable decision. The purpose of our research is to improve the use of visualization in maritime surveillance systems, which mostly handle unexpected events. In the rest of this paper, we focus on unexpected events within control systems.

Reference [21] defines control as "the ability to direct and manage the development of events, and especially to compensate for disturbances and disruptions in a timely and effective manner" (p. 148). Collaboration between human and machine in cognitive systems for control should allow measuring and interpreting differences between actual and "intended" states [21]. This way, unexpected events would be defined by the observed offset. If this offset is known, users would understand *why* this situation was not expected and *how* to deal with it.

In order to maintain control, Hollnagel proposed the Extended Control Model (ECOM) [5]. This is a four layer control model for risks assessments, instead of a basic control loop: (1) characterise the context, (2) identify the risks, (3) analyse the risks and (4) decide on the countermeasures. As each control layer affects the lower one, risks identification would depend on the context, analysis would depend on the possible risks, etc.

Within the context of maritime surveillance, Idiri and Napoli proposed a new definition of risk that takes into account geographical and behavioural specification [28]: it is the combination of a vessel's behaviour (based on its kinematic), a geographic area (e.g., dangerous zone or not) and a situation (e.g., vessel type, visibility, meteorology).

We observed that MSS do not take into account these three elements, and are not based on the ECOM model for controlling the maritime domain, though these control tools use both human and machine contribution. Moreover, previous work in visual analytics for maritime surveillance usually proposed only one or a few specific methods for traffic visual analysis.

Various visual analytics methods have been identified to answer ECOM control loop, allowing visual analysis at different scales and for various tasks. Choropleth maps, density maps, statistical analysis and clustering methods provide cartographic analysis methods for characterising geographic areas [12], [29]: at sea, we can visualize zones with high density of vessels, risky areas (piracy, drug traffic). Figure 3 is an example of clustered data, using the same ships positions than Figure 1. Results are obtained with an OpenLayers map using a Cluster strategy. Another type of area characterisation is displayed on Figure 4, giving an example of heat map that shows major traffic zones in the Mediterranean Sea.


Figure 3. Cluster method to display ships position with an OpenLayers map (FishEye, MINES ParisTech - CRC).

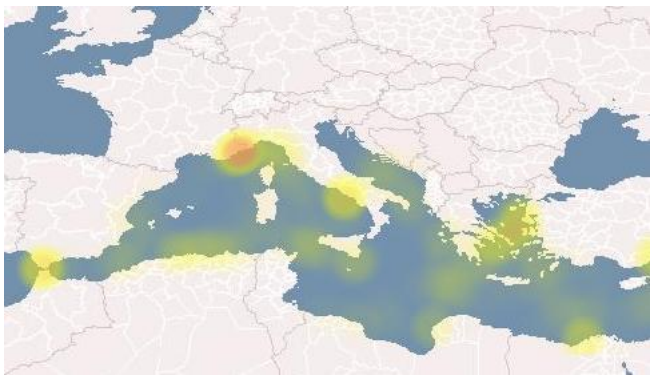


Figure 4. A heat map shows zones with high density of vessel (FishEye, MINES ParisTech - CRC).

Visualizing a vessel's profile with parallel coordinates plots (e.g., Figure 5), speed profile or past trajectory, and comparing it to other vessels profile are examples of support for the controller's work in identifying and analysing risks. Depending on the user's profile and education, advanced

 ISBN: 978-1-61208-251-6 

visualization environments can be used: from simple diagrams to highly interactive 3D visualization in a space-time cube. Figure 5 uses PCP visualization to describe ships profile with five quantitative attributes: ship length, ship draught (distance between waterline and bottom of hull), longitude, latitude and speed (in knots). Vessels data are provided in near real-time by DCNS from AIS sensors. The colour of the lines corresponds to the country attribute of the ships. As this sixth attribute is qualitative information, using colour variable is preferred to creating a new axis.

The selectivity of the colour variable is the easiest mean to detect particular profiles on this type of diagram. As an example, we clearly distinguish a correlation between the nationality of ships and the geographic coordinates (axis 3 and 4) on Figure 5. Purple lines stand for Italy (mean coordinates: N41° E012°), green for Greece (mean coordinates N38° E025°), blue for France (mean coordinates N43° E004°) and orange for Spain (mean coordinates N40° E002°).

The colour variable could also represent an attribute such as the type of ship, or its position within a previous classification. This will give more information in a simple line diagram that could be used by controllers or analysts.

To respect the philosophy of visual analytics, this type of diagram must be highly interactive: brushing tools allow the selection of a subset of data, and the map is synchronised to the selection. This way, both the geographic and the attributes dimension are taken into account [23].

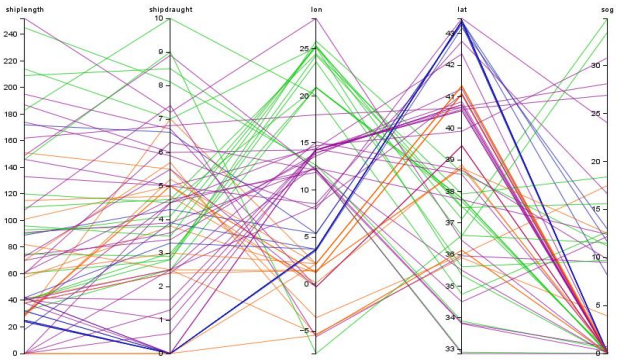



Figure 5. Parallel coordinates plot describing the profile of 70 ships (ship length, ship draught, lon, lat, speed over ground).

Cognitive systems for the control require as many geovisual analytics methods as there are maritime risks and users for the system. We observed that proposing a single method that only suits a single need cannot fulfil all the steps that were described in this section. Geovisual analytics contribution, evaluation and user's needs have to be investigated first [13].

Therefore, the use of geovisual analytics strongly depends on the type of data to analyse, the user in charge of the task and the purpose of the study. In the next section, we describe the approach that will be used in upcoming research work.

## IV. GUIDING THE USER IN VISUAL ANALYTICS

The philosophy of geovisualization and geovisual analytics is to use both maps and diagrams to explore and analyse data [23]. As it was presented in Section II, many different methods can be applied to analyse traffic data. Moreover, cognitive systems for control highlight various situations that have to be taken into account in risk control, depending on available data, users and questions to be answered.

Using several methods of geovisual analytics involves guiding the user in the choice of these ones. As analysts or controllers do not have the same knowledge about data exploration than a computer scientist, an automatic process based on visualization and risk knowledge should help the user in this analytical task.

We propose to formalize (1) the needs in visual analytics for maritime risk management and (2) the contribution of visual analytics methods to data analysis. Knowledge bases would be the support for automatic proposal of methods or tools to the user.

Figure 6 explains the architecture of such a methodology, using a knowledge-based system to propose the most suitable methods to be applied, according to the user and the tasks to perform. The input of the system are the user, who has a specific pre-defined profile (controller, analytics, researcher, etc.), the available data (characterised by their type, their amount, their geographical extent and statistical information) and eventually the tasks that have to be performed.

The Risk Management base will be used to extract useful tasks to complete the input data, according to the user's profile. As an example, if the user logged as the captain of a vessel, the first tasks that should be perform are answering the questions "Where am I?", "What is the situation around me?", "Is there any risk of collision?".

Using the evaluation and characterisation of geovisual analytics methods from the Visual Analytics base, the inference engine will process the information to evaluate the most suitable tools: its means the most effective, useful and usable tools according to the user, tasks and data. With the same example of a captain, proposed results could be: an overview map centred on the vessel's position, a relative view of surrounding vessels, his own course and a potential risk map evaluating distance from other ships.

The user interface is the most important part of the process, as it would be the main way to exchange information between the user and the process: it should allow easily changing the profile or changing the tasks to perform; and therefore change the proposed visual analytics methods in the output.

A knowledge-based method was proposed by Beaulieu [30] to suggest the most suitable representation models for multivariate data in a SOLAP system [31], which allows exploring and analysing massive heterogeneous datasets. This research was based on the visual variables of Bertin [32] within a "semiological knowledge base": rules based on

 ISBN: 978-1-61208-251-6 

the semiology of graphics process the input data, returning possible types of graphic display. This way, representation models such as the type of diagram, shapes and visual variable(s) were proposed in order to respect graphic semiology, according to the specificity of data set and user's request.

Proposing a knowledge-based system to guide the user in visual analytics for risk management allows keeping the diversity of geovisual analytics environments, without developing a unique visualization mode. Depending on user's input (profile, purpose, available data), an overall view of possible methods will suggest to the user the most suitable ones. This overview of visual analytics environments allows a high interactivity level between users and data: this is the founding principle of visual analytics [17], [33].

Our following research will focus on characterising and evaluating specific visual analytics methods and tools for risk management, taking into account two aspects: their specificities of use (e.g., input, output data, design, limits, processing time) and the user's perception (e.g., ease of use, usefulness, time of use, types of queries).

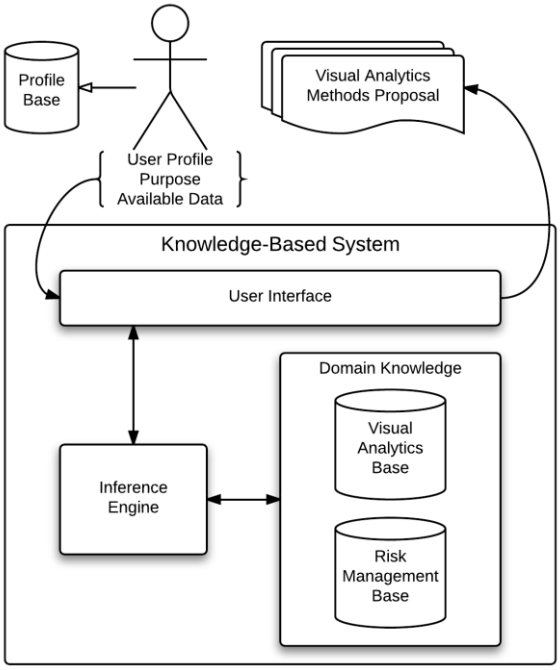



Figure 6. Architecture for a guided selection of visual analytics methods. Based on selected user profile and situation inputs, an knowledge-based system would suggest the most suitable visual analytics methods.

## V. CONCLUSION AND FUTURE WORK

In this paper, we showed that geovisual analytics methods provide efficient results to explore or analyse massive and multidimensional data, and are still missing within maritime surveillance systems. But their diversity is also a limit when considering all possible users and uses. Knowing which type of visualization or which analysis has to be led requires a real knowledge of these methods and how they have to be used.

Operators in maritime surveillance are faced to many heterogeneous data to be controlled in real-time; analysts investigate past events to extract new knowledge; captains confront facts "in the field" and do not perceive the information the same way operators do. Each one of these profiles requires its own method of data visualization and visual analytics in order to have the best situation awareness.

In order to improve modelling, visualizing and analysing maritime information, we have seen that a single visual analytics method does not match the diversity of these needs. The basis for a methodology for selecting proper geovisual analytics methods was introduced, based on users needs in risk control and visual analytics expertise. With the use of these two knowledge bases, the best visual analytics methods that fulfil one's requirements would be suggested.

We introduced a new approach for considering user's profile and purposes in maritime surveillance systems, based on the evaluation of visual analytics environments. This evaluation takes into account the specificities of each visualization method and its perceived usefulness and ease of use.

Geovisual analytics would both use computer methods to summarize data and user's knowledge in data exploration. Even if cognitive systems for the control are be limited by human factors and human errors, they are major steps in controlling and decision making.

To develop the method that was presented in this paper, we plan to study risk management requirements in analysis of data. Then, this study will be applied to maritime domain with our research partners in maritime surveillance. Geovisualization and information visualization methods will be tested and evaluated with people who are used to work with maritime data, but who have various profiles and different knowledge in information visualization.

## REFERENCES

[1] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems", Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 37, no. 1, p. 32-64, 1995.

[2] J. Noyes and M. Bransby, Ed., People in Control: Human factors in control room design. The Institution of Engineering and Technology, Michael Faraday House, Six Hills Way, Stevenage SG1 2AY, UK: IET, 2001..

[3] S. I. Fabrikant and A. Lobben, "Introduction: Cognitive issues in geographic information visualization", Cartographica: The International Journal for Geographic Information and Geovisualization, vol. 44, no. 3, pp. 139-143, 2009.

[4] R. Parasuraman, T. B. Sheridan and C. D. Wickens, "A model for types and levels of human interaction with automation", IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 30, no. 3, pp. 286-297, May 2000.

[5] E. Hollnagel, Barriers and Accident Prevention. Ashgate Publishing Limited, 2004.

 ISBN: 978-1-61208-251-6 

[6] A. Vandecasteele, A. Napoli and M. Morel, "Apport du géodécisionnel et de la géocollaboration aux systèmes de surveillance maritime", presented at INFORSID 2011, Lille, France, 2011.

[7] V. Mallé-Noyon, "Modèle de vue relative et spatio-temporelle de trajectoires géographiques d'objets mobiles : Application au contexte maritime", CFC, no. 198, December 2008.

[8] W. van Hage, V. Malaisé, G. de Vries, G. Schreiber and M. van Someren, "Abstracting and reasoning over ship trajectories and web data with the simple event model (SEM)", Multimedia Tools and Applications, p. 1-23, 2011.

[9] B. Idiri and A. Napoli, "Découverte de règles d'association pour l'aide à la prévision des accidents maritimes", in Revue des Nouvelles Technologies de l'Information, Bordeaux, France, 2012, vol. 23, pp. 243-248.

[10] X. Chaze, A. Bouejla, A. Napoli and F. Guarnieri, "Integration of a Bayesian network for response planning in a maritime miracy risk management system", Proc. 7th International Conference on System of Systems Engineering (SOSE 2012), Genoa, Italy, 2012.

[11] S. Fournier, "Intégration de la dimension spatiale au sein d'un modèle multi-agents à base de rôles pour la simulation: application à la navigation maritime", Université de Rennes 1, 2005.

[12] N. Willems, "Visualization of vessel traffic", Technische Universiteit Eindhoven, 2011.

[13] M. Riveiro, "Visual analytics for maritime anomaly detection", University of Skövde, 2011.

[14] M. Morel, A. Napoli, J.-P. George, F. Jangal, M.-A. Giraud and M. Botalla, "Surveillance et contrôle des activités des navires en mer : ScanMaris", Proc. 4th Workshop Interdisciplinaire sur la Sécurité Globale (WISG 2010), Troyes, France, 2010.

[15] A. M. MacEachren and I. Brewer, "Developing a conceptual framework for visually-enabled geocollaboration", International Journal of Geographical Information Science, vol. 18, no. 1, pp. 1-34, 2004.

[16] G. Cai, "Extending distributed GIS to support geo-collaborative crisis management", Geographic Information Science, vol. 11, no. 1, pp. 4-14, June 2005.

[17] J. J. Thomas and K. A. Cook, Éd., Illuminating the Path: The Research and Development Agenda for Visual Analytics. National Visualization and Analytics, 2005.

[18] B. M. Tomaszewski, A. C. Robinson, C. Weaver, M. Stryker and A. M. MacEachren, "Geovisual analytics and crisis management", Proc. 4th International ISCRAM Conference, Delft, Netherlands, 2007.

[19] M. Riveiro and G. Falkman, "The role of visualization and interaction in maritime anomaly detection", in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 2011, vol. 7868, p. 20.

[20] D. C. Foyle, A. D. Andre and B. L. Hooey, "Situation awareness in an augmented reality cockpit: Design, viewpoints and cognitive glue", Proc. 11th International Conference on Human Computer Interaction, Las Vegas, NV, 2005.

[21] E. Hollnagel and D. D. Woods, Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. CRC Press, 2005.

[22] G. Andrienko, N. Andrienko, J. Dykes, S. I. Fabrikant and M. Wachowicz, "Geovisualization of dynamics, movement and change: Key issues and developing approaches in visualization research", Information Visualization, vol. 7, no. 3-4, pp. 173-180, 2008.

[23] M.-J. Kraak, "Beyond geovisualization", Computer Graphics and Applications, IEEE, vol. 26, no. 4, pp. 6-9, August 2006.

[24] J. Chen, A. M. MacEachren and D. Guo, "Supporting the process of exploring and interpreting space–time multivariate patterns: the Visual Inquiry Toolkit", Cartography and Geographic Information Science, vol. 35, no. 1, p. 33, 2008.

[25] H. Guo, Z. Wang, B. Yu, H. Zhao and X. Yuan, "TripVista: Triple Perspective Visual Trajectory Analytics and its application on microscopic traffic data at a road intersection", Proc. Pacific Visualization Symposium (PacificVis 2011), IEEE, March 2011, pp. 163-170.

[26] M. Davenport and C. Risley, "Information visualization: the state of the art for maritime domain awareness", DTIC Document, 2006.

[27] J.-L. Wybo, Maîtrise des Risques et Prévention des Crises, Lavoisier, in press.

[28] B. Idiri and A. Napoli, "Towards automatic identification system of maritime risk accidents by rule-based reasoning knowledge", Proc. 7th International Conference on System of Systems Engineering (SOSE 2012), Genoa, Italy, 2012.

[29] D. Pumain and T. Saint-Julien, L'analyse spatiale : Localisations dans l'espace. Paris: Armand Colin, 1997.

[30] V. Beaulieu, "Étude de la visualisation géographique dans un environnement d'exploration interactive de données géodécisionnelles - Adaptation et améliorations", Université de Laval, 2009.

[31] Y. Bédard, S. Rivest and M.-J. Proulx, "Spatial On-Line Analytical Processing (SOLAP): Concepts, architectures and solutions from a geomatics engineering perspective", in Data Warehouses and OLAP: Concepts, Architecture and Solutions, Wrembel R. et C. Koncilia (Eds), 2007, pp. 298-319.

[32] J. Bertin, Sémiologie Graphique, Mouton/Gauthier-Villars. Paris, 1967.

[33] D. A. Keim, F. Mansmann, J. Schneidewind and H. Ziegler, "Challenges in visual data analysis", Proc. 10th International Conference on Information Visualization, London, UK, 2006.

 ISBN: 978-1-61208-251-6 

# Digitizing the Geo-Cover of Greater Cairo Metropolis Supporting the Time of Transformation

Ahmed Abdelhalim M. Hassan
Institute of Landscape Ecology, Muenster University,
Muenster, Germany,
E-mail ahmedahalim@uni-muenster.de

***Abstract*—Nowadays, many cities in developing countries are striving for reconstruction and sustainable growth after a long time of corruption and conflict. Hence, the planners of different trends need to access different types of information easily and work firmly by the use of modern technology. Accordingly, this paper aims to present the geological and physiographic setting of Greater Cairo metropolis in the perspective of a digital city and reintroduce the geo-map concern to the whole area. To achieve that, this study tends to use the remote sensing techniques to differentiate the rock cover types and describe the topographic relief of the area dominated Greater Cairo. The applied classification and nomenclatures were based on field observations that take the previous geological studies published on the concerned area and its neighborhoods into strong consideration. As a general result, eleven units representing all rock outcrops have been observed, described, and classified. Analyzing of the digital elevation model (DEM) delineated the structure relief of the study area. These outputs could support the stakeholders and earth-scientists to refine and develop a geological data bank in a digital form, to be directly processed by suitable software. And also, the familiarity with the geo-setting of such a megacity would help to detect the rapid spatial changes in some protected areas.**

***Keywords-digital geo-cover; Greater Cairo physiography; rock unit classification***

## I. INTRODUCTION

Generally, applications of techniques to obtain information about an object without touching the object itself are used widely and extensively in geological investigations. These techniques mainly are geophysical tools (e.g. electromagnetic induction, ground penetration radar, aeromagnetic…etc.) and satellite sensors. These were located far apart hence the distance between the object and sensor being of several kilometers or hundreds of kilometers [1]. In the recent past, the remote sensing technique plays a very important role in geological mapping starting from aerial photographs interpretation and jumping to sophisticated enhancement, processing, and interpretation of images acquired by space satellites. These are able to show features and patterns which may not be distinguishable on aerial photographs due to the lack of color information in these (aerial photographs) [2]. Therefore, the remote sensing techniques used do not only allow geological mapping but also lithology and mineral differentiation and exploration on small and large scales, respectively. Accordingly, the use of satellite images for geological mapping and for exploring economic resources is becoming an increasingly important issue for earth-scientists.

Furthermore, the 3D or elevation remote acquisition data provide mitigation and hydrological investigation for geo-hazards rather than topographic description and relief description and measures. The later proved that the application of digital elevation model analysis (DEM) is a potentially efficient, reliable, reproducible and effective technique for under taking geological terrain mapping. Abd Manap et al. [3] concluded that the advantage of 3-D draping technique compared to the conventional stereoscope interpretation is that the geological terrain features such as hillcrest, side slope, foot slope, straight slope, concave slope and convex slope can be observed not only from the normal vertical view but also to be viewed from different scales, orientations and perspectives.

A Study done by Nalbant and Alpiekin [4] showed that thematic mapper (TM) imagery can be used as a valuable tool together with field studies for geological mapping and structural patterns. This procedure can save an appreciable amount of time, money, and man power compared to the efforts exerted by earth scientists and is also useful in undertaking geological terrain mapping in inaccessible areas. The U.S Geological Survey carried out a research program in 1985 to produce 1:250,000 scale land-cover maps for Alaska using Landsat MSS data [5]. Sultan et al. [6] and, Gad and Kusky [7] demonstrated that TM data can be reliably used to distinguish mineral potential from surrounding rocks in arid regions and to generate detailed maps over wide regions by using quantitative, reproducible mapping criteria. In addition, possibilities for locating suture zones over the less well known parts of arid continents are clear.

Ingram et al. [8] evaluated the relationships between the geology, land use, and elevation parameters in north Mississippi and addressed a strong correlation between the different rock formations and slope degrees. Since, land use information was extracted from satellite imagery, topographic parameters were derived from elevation data, and textural characteristics were generated from these datasets to provide a basis for surface mapping. However, an integrated GIS (geographic information system) and remote sensing techniques can be used effectively to develop a more comprehensive geologic database to facilitate geological field studies have been occurring for large areas.

 ISBN: 978-1-61208-251-6 

The most geological studies describing Greater Cairo lack the application of techniques of remote sensing in an advanced way. Hence, the most work deals with raw image data by traditionally using color combination of multispectral satellite images to prepare preliminary maps regardless of the georeference, radiometric character or spectral analysis. Also, the most maps were traced based on Thematic Mapper (TM) and Enhanced Thematic Mapper (ETM) which are recently available freely in many open sources. Such studies conducted by Egyptian geologists (e.g., [9]) who modified the geologic map of Eastern Greater Cairo, area based on false color composite and utilization of band combination 7, 4, and 2 for the R, G, and B, respectively, which is the best for the general lithological discrimination in the study area.

## II. T STUDY AREAYPE

The selected area of study is the metropolitan area of Greater Cairo (G.C) and its surroundings, which is known as the capital of Egypt and one of the fastest growing megacities worldwide (Fig. 1). The area covers about 600 km2, encompassing major parts of the governorates of Cairo, Giza, 6th of October, and Helwan. The Nile forms the administrative division between these governorates, with Cairo and Helwan on the east bank of the river and, Giza and 6th of October on the west bank. Also, the area is geologically divided into two main parts, eastern part and western part, with the Nile valley between them. In relation to the later, the nomenclature and facies subdivision of the stratigraphic succession, described by many authors on both sides, are quite different. Therefore, the rock type characters, classified in this study, have been implemented on the identification and chronology of previous studies.

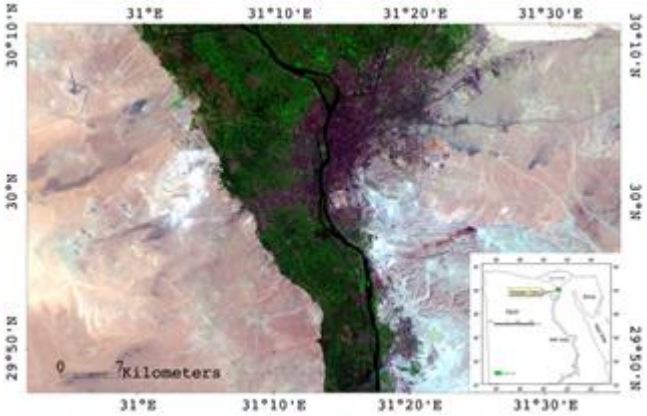



Figure 1. Location of the study area.

## III. MATERIALS AND TECHNIQUES

### A. Data availability

Because of the rapid development and extensive human activities in the area around the Nile valley in Greater Cairo (hinterland and/or Nile valley terraces) the selected data are characterized by less cover of land use for a better discrimination of the rock units. Accordingly, Landsat 5 TM image data acquired in 1984 with seven bands of (three invisible wavelengths, four in infrared) most of which have 30 meter resolution, except band six with 120 m resolution, has been selected. Then to increase the spatial accuracy of TM-5 image the moderate resolution panchromatic image with 10m pixel resolution acquired in 1997 by Spot-2 sensor (TT station) is merged with it. The high resolution (5 m pixel resolution) of multispectral Spot-5 images acquired in 2006 by Spot-5 is available too for a good and accurate glance of surface exposures in the study area. Moreover, there is another sort of satellite data with elevation information that has been used to construct 3D visualization for the entire area of interest. This elevation data is obtained from ASTER GDEM which is generally characterized by 30 m pixel resolution at 95 % confidence horizontally and about 10 m at 95 % confidence vertically. Furthermore, ancillary data such as scanned topographic sheets and geologic map of scale 1:100,000 are utilized for required nomenclatures and location detection.

### B. Methodology

In this study, the work procedure is based on different techniques of image processing to extract some interesting geologic and physiographic features. First of all, all satellite images and scanned sheets are georeferenced to UTM Zone 36 North projection with WGS-84 datum. Moreover, TM and Spot satellite data are enhanced to insure radiometric balance between individual scenes and the DEM data were mosaicked to cover the study area. This step was followed by fusion technique to increase the pixel size and resolution of all TM image bands (Fig. 2) to get more spatial information of surface exposures. However, the produced maps based on TM reached to scale 1:100,000, but the maps based on merged image reached to scale 1:32,000 without remarkable pixels.

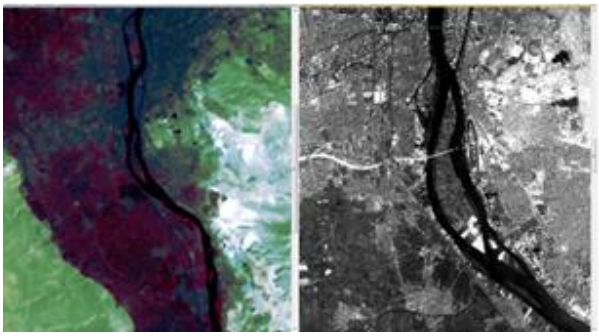


Figure 2. TM image (1984) to left and Panchromatic Spot band (1997) to right represent low and high resolutions respectively.

The fusion process occurred by using the sharpen module which is provided with different merge algorithms in Erdas Imagine software. Therefore, the subtractive resolution merge algorithm is used in this study. Consequently, the different band combinations are tested for first impression on lithology discrimination. So, the combination between bands 7, 4, and 2, and 5, 3, and 2 were reviewed in R.G.B false

 ISBN: 978-1-61208-251-6 

color and presented the best image composites in the study area.

The principal component analysis used in this study to reproduce the merged high resolution image allows inordinate and excessive data to be computed into fewer bands [10]. In addition, Principal component algorithms are image enhancement techniques to visualize the maximum spectral contrast from many spectral bands to three primary display colors [11]. These lead to a stretching of the pixels to differentiate different rock types. Moreover, the bands of PCA data are non-correlated and independent and are often more interpretable than the source data [12]. The PCA technique was applied onto two different kinds of multispectral satellites data TM with 7 bands and Spot with 3 bands separately. The difference after the removal of redundancy information showed that differences exist between the different bands of each image. Because of the wide range of spectral wavelength of TM data which have seven channels, the author preferred to use principle component analysis of TM merged image in rock discrimination to compose a geo-digital thematic map rather than the multi-spectral Spot data which have only three channels.

Consequently, it could found that the first principal component channel (PC1) has the largest possible variance, while the next two PC channels (PC2 and PC3) contain all other interband variations. Therefore, each one of the components from 3 to 7 in TM merged image is dominated in less than 1.3% of information and seems unnecessary in lithological information (Fig. 3).

So that, the first three principle components, these were calculated from preferred and enhanced TM merged image beside the field observations and experiences, have been mainly used in classification process.

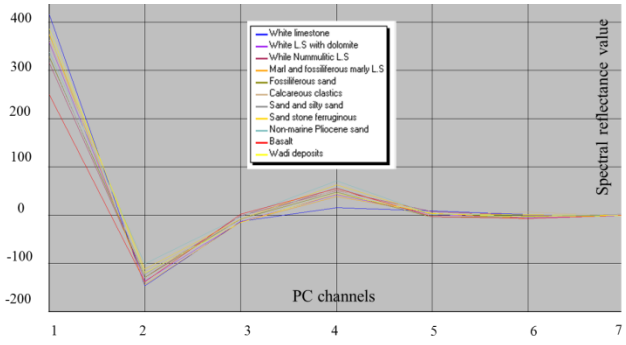



Figure 3. The spectral reflecance of the classifies rock typs regardind the PC channels.

The classification step is provided by the supervised classification method which assembles the surfaces that have similar spectral signatures. Once signatures are examined, the classifier is then used to attach labels to all image pixels according to the trained classes. The procedure of supervised classification was processed by using maximum likelihood algorithm, which assumes that each spectral class can be described by a multivariate normal distribution [13]. In spite of that, the accuracy assessment is behind the objective of this study; but, there are many control points acquired from ground and estimated by GPS used and interpolated in the process of classification to enhance the pixel selection and refine the rock type identification.

Finally, the morphological information and topographic features were extracted from the available mosaicked digital elevation model (DEM) (Fig. 4).

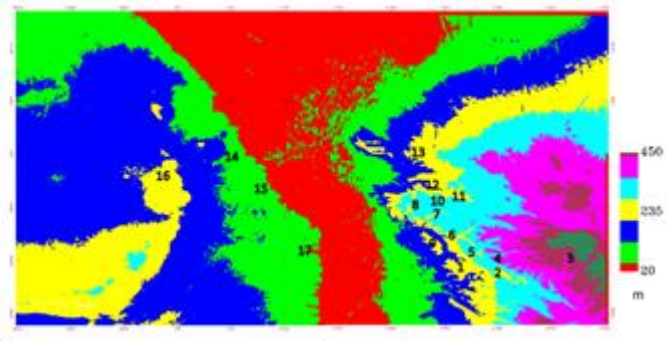



Figure 4. Colorued DEM for elevation information concerns the Greater Cairo area.

1-15th of May City. 2- Wadi Garawi. 3- El Halawana Height.
4- El Qurn Height.. 5- Wadi Gibbu.. 6- Wadi Abu Silli.
7- Wadi Hof. 8- Gabal Hof 9- Obesrvatiory Plateou.
10- Wadi Abu El Rokham. 11- Wadi Degla. 12-.Wadi El Tih.
13- El Mokattam Plateou. 14- Abu Roash. 15- Pyramids Plateou.
16- 6 th of October City.17- Wadi El Tafla.

The techniques used to obtain and analyze the elevation information are mainly supported by ENVI 4.7 software. As well, the produced digital maps could be reached to scale from 1:75,000 to 1:50,000.

## IV. GEOLOGY OF GREATER CAIRO

### *A. Rock Cover and Type*

Many authors studied the different time rock units cropping out in the eastern and western sides of Greater Cairo and reported that in geographic maps and geologic correlations e.g., [9][14][15][16][17][18][19][20][21].

The present application of remote sensing image processing techniques and field observation supplemented with the previous geological studies and nomenclatures in West-and-East Greater Cairo came to subdivide the rock exposure of the study area into the eleven rock types. Besides that, two dominating classes of the fertile Nile valley area are concerned too. One of them represents the urban materials and roads, and other one is concerned as a mix class representing the green cover and water bodies.

Moreover, the spectral analysis of the exposed rocks in the area of interest is summarized and reported to show the relation between each band of PC image and reflectance spectrum of selected pixel for classification. (Fig. 3).

In short, the eleven litho-types covering the study area are mainly dominated by 1-white limestone (Chalk), 2- white limestone with dolomite, 3- white Nummulitic limestone, 4- marl and fossiliferous marly limestone, 5- sands and fossiliferous sandstone, 6- calcareous sandstone and/or

 ISBN: 978-1-61208-251-6 

eroded surface 7- sand and silty sand with subordinate clay intebeds, 8- sand stone (Ferruginous) with Gravels, 9- non-marine Pliocene sand, 10- basalt, and 11- wadi deposits (Fig. 5).

These exposed rock succession that builds up the study area ranges in age from Upper Cretaceous to Quaternary (Fig. 6).

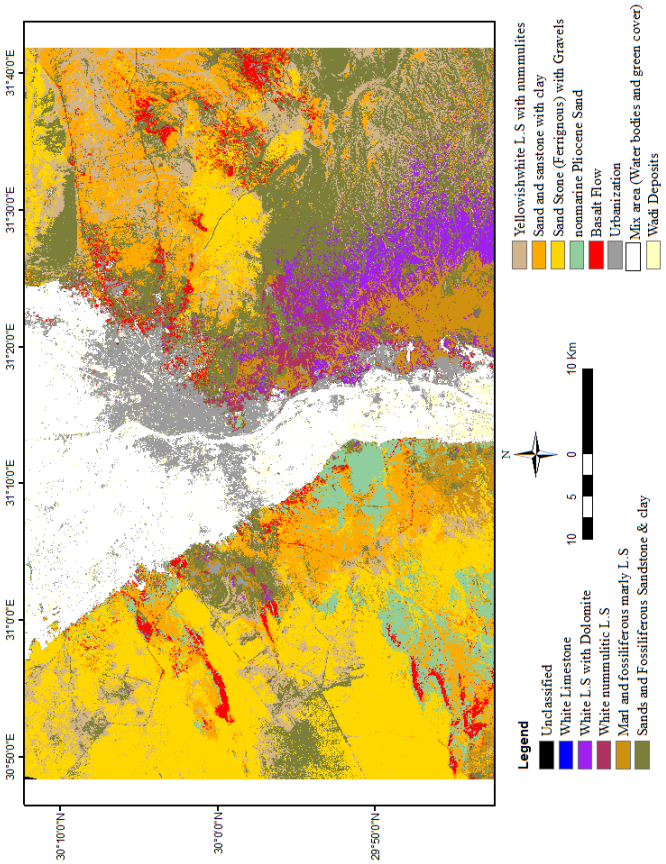



Figure 5. Rock cover classification based on PCA.

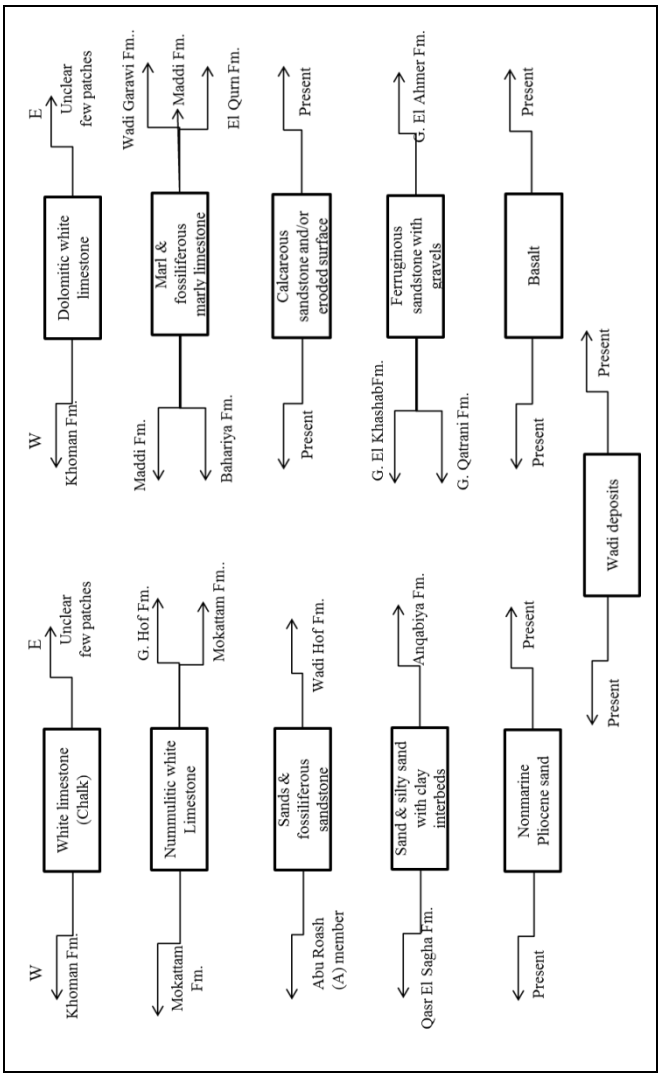



Figure 6. Rock type and its equivalent formation

*1) White limestone (L.S)*

White limestone or chalk unit is distinguished by the first appearance of snow-white massive thick bedded chalk, with thin chert bands and nodules (Fig. 7 Left). The chalk is exposed only in the Western part of the study area mainly in Hassana Dom. According to Said [15], the age of Chalk unit ranges from Campanian to Masstrichtian with a maximum exposed thickness of about 78 m. While in western part of the desert this formation belongs to the Khoman Chalk and Senonian age [22].

*2) White limestone with dolomite*

The white limestone contaminated with dolomite and / or marl is represented in the eastern part of the study area mainly by Observatory Formation and slightly by Abu Roash Formation or AbuRoash "D" member in Western part (Fig. 7 Right). Furthermore, the Observatory Formation belongs to the Middle Eocene and constitutes the foundation bedrock of the northern eastern part of 15th of May city and its northern

 ISBN: 978-1-61208-251-6 

extension [9]. Swedan [20] mentioned that the formation is composed of white to yellowish white, marly and chalky limestone, intercalated with several interbeds of hard, grey, dolomitic limestone. The type section of this rock unit measured by Frag and Ismail [14] in the northeast of Helwan below the Observatory establishment is about 77 m thick.

On the other side, the Abu Roach Formation or "D" member [16] is dominated mainly by white to yellowish white, dolomitic limestone with large amounts of large fauna (e.g. *Acteonella*). [15], named the Turonian-Santonian beds in many parts in northern Egypt (included the study area) as Wata Formation, commonly used in Sinai.

*3) White Nummulitic limestone*

The white Nummulitic limestone class refers to the beds built up mostly with grayish-white, slightly chalky beds rich in *Nummulites gizehensis, Nummulites Beaumonti, N. subbeaumonti, Schizaster africanus*, and *Turbinella frequens* [14]. These beds exposed in the Eastern part to the North of Helwan province which is composed by upper part of the Gabal Hof Formation. The former named and subdivided into two units (up to 80 m in thickness) in its type locality at Wadi Abu Rakham by Frag and Ismail [14]. This formation is equivalent to the Middle Eocene rock unit cropped out at the escarpment of El Mokattam Plateau, East of Cairo city. While, along the western side of the Nile, the spectral analysis of this rock type does not show any good and/or obvious cover except at the area of Pyramid plateau.

*4) Marl and fossiliferous marly limestone*

The marl and/or fossiliferous marly limestone are mainly represented in the eastern side of the study by Upper-Middle to Upper Eocene time unit of El-Quran and Wadi Garawi Formations in southeast part and EL-Maddi Formation in northwest part. Osman [9], Frag and Ismail [14], Swedan [20] and we mentioned that the El Qurn Formation is composed mainly of about 70 m thick marl, marly limestone with thick gypsum veinlets, and hard dolomitic bands. In addition, they stated that the Wadi Garawi Formation is extremely equivalent to the Maddi Formation, which dominated by *Carolia Placunoides, Plicatula polymorpha, Ostrea reili, Nummulites Beaumontai*, and *N. striatus.*

On the West side, this rock class is recorded southwest and west from the Pyramid plateau with maximum thickness of 154 m measured by Frag and Ismail [14] and named Maddi Formation. To the north of the Pyramid plateau, this unit cropped out at the flanks of macro structure of El Hassana Dom. In contrast, the spectral signature of this class has referred to the marl and fossiliferous marly limestone units which intercalated with sand and shale in the small part of the core of Abu Roash structure (Fig. 8). These units belong to the Cenomania (Lower Cretaceous) Baharyia Formation [15].

*5) Sands and fossilifreous sandstone*

The sand stone facies cover a large area east and west of Greater Cairo, which constitute the foundation bedrock of many parts in new cities such as New Cairo, and 6$^{th}$ of October City. This rock type is composed mainly of sand stone, marl, yellowish green clayey marls, sandy limestone and bioturbated marly limestone. According to the describtion of Frag and Ismail [14], Moustafa et al. [18] and Swedan [20] the spectral signature of this class could be represented by Late Upper Eocene Wadi Hof Formation and Anqabiya Formation in eastern side of Nile valley. While, on the other side this reflectance spectrum belongs to the upper part of Qasr El Sagha Formation. This formation concerned by Said [15] and Swedan [20]. The upper most part is of the Eocene age and consists of poorly fossiliferous clay and channel deposits with minimum thickness regarding the original thickness. Also, this signature reflected the Coniacian-Santonian beds exposed in some areas at El-Hassana Dom.

*6) Calcareous clastics and/or eroded surface*

This spectral reflectance is very ambiguous all over the study area, but the field check oriented to identify it to be under the erosional surfaces of clastics and calcareous-clastics outcrops, in addition to some eroded Nummulitic limestone with clay contamination.

*7) Sand and silty sand with clay*

The exposed part of this rock type is found mainly at Gabal el Mokattam area eastwards and very close to the area west to southwest of The Pyramids Plateau composed mainly of fine grained calcareous sand stone, grading upward silty sand and shale [20] and [9]. The unit could be represented in the Anqabiya Formation due East of the study area and some most upper Eocene and Oligocene units due West (Qasr El sagha Formation). Generally, this pixel reflectance represents the unconsolidated sandstone and silt to shale outcrops in many parts of the study area.

*8) Sandstone (ferruginous) with gravels*

This exposure is widely distributed in the study area, especially East of el Mokataum plateau including the East of Cairo (Nasr City), New Cairo City, and protected areas (forest of petrified wood) along the Cairo-Red Sea high way. This rock exposure belongs to the Gabal Ahmar Formation. On other hand, this rock cover mainly dominates the region of 6$^{th}$ of October City and belongs to Gabal Qatrani Formation and Gabal El Khashab Formation [20]. However, Skukri [23] and Said [15] stated that, this formation constitutes the extension of a narrow belt of Oligocene-Miocene age from Suez via Cairo and onward into the north western desert, which is mainly dominated by distinct red bed sequences with coarse-grained sand and gravels with large amounts of petrified wood separated by basaltic sheets in between (Fig. 9 Left). The pixel signature information is reported as following;

*9) Non-marine Pliocene sand*

This rock unit in the area of interest consists mainly of fine to coarse, friable sands and conglomerate sandstone of yellowish to brownish-white color [15] and [20]. Its spectral reflection appears on the western side of the G.C obviously accompanied with basaltic sheets rather than on the eastern side (Fig. 9 Right).

 ISBN: 978-1-61208-251-6 

*10) Basalt*

The distribution of Oligo-Miocene deposits was governed by the volcanicity and tectonic activity which affected the Red Sea regions and the high belt between the stable and the unstable shelves in Egypt during the Oligocene [15]. Therefore, the basaltic sheets in the study area recorded within the sandstone beds of Oligo-Miocene beds in 6th of October city towards the West and New Cairo area towards the East.

*11) Wadi deposits*

In the study area which is filled with recent fine clastics deposits mainly of clay, this rock unit represents its often being reworked by eolian processes and stream deposits near to the River Nile bank.

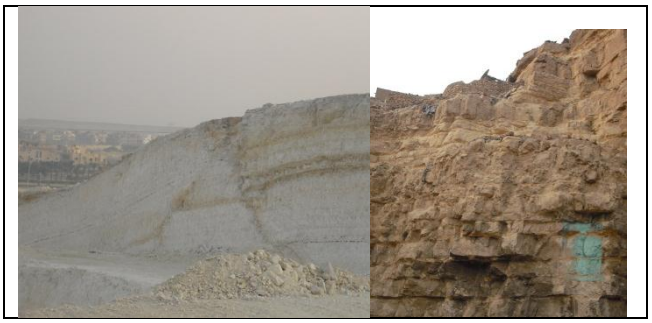


Figure 7. Chalky limestone (Khoman Fm.) observed in the western part of study area (Left). Well bedded limestone at Gabl Hof Fm. (Right).


Figure 8. The core of Abu Roasch structure shows the marl and limestone rock units.

Figure 9. Petrified wood at Gabal El Khashab, West of G.C (Left). Non-marine Pliocene caped by basaltic sheet (Right).

*B. Physiography*

Topographically, the study area could be divided into three parts, the first one is represented by the lowest elevation corridor of the River Nile valley, the second is dominated by two plateaus detached by two low areas at the eastern side of the area under investigation, while, the last is the western side which represented by topographic features forms an almost featureless plain with the exception of some small structure related heights

According to the analyzing and slicing of the digital elevation model (DEM), the contour lines and elevation information, slope degradation, and three dimensional visualization are extracted (Figs. 4 and 11).

Therefore, the area around River Nile mainly occupied by different human activates such as agricultures, houses, and industries with elevations ranging from 20 to 100 m above sea level.

The topographic features expressed on the East side of G.C characterized by moderately rough relief, occupied by soft rocks ranging in age from Middle Eocene to Recent.

The highest elevation is about 450 m above sea level (a. s. l) recorded due East of the El Halawana Hieght. In general, the landscape in the part of study area is characterized by numerous rugged, isolated hills of mostly made up of hard Eocene limestone beds. Those are arranged from south (close to 15th of May City) to north (El Mokattam area), e.g. El Qurn Hieght (~270 m a.s.l), Obesrvatory table land (~175 m), Gabal Hof (~330 m a.s.l), and El Mokattam Plateau and New Cairo area (maximum ~287 m a.s.l). The wadies are usually controlled by faults of different trends mainly of NE to N-S and E-W to ENE trends. The main wadies traversing the study area are Wadi Garwi to the southeast from 15th of May City, Wadi Gibbu and wadi Abu Silli which dissect the 15th of May City and Helwan area towards due West, to North of Helwan area present Wadi Hof which is extending ESE-WNW and joined with Wadi Abu El Rakham at Gabal Hof area, Wadi Degla which is the most important protected area and largest drainage line on the east side of G.C delineated the southern scarp of El Mokattam plateau, and due North of the Wadi Degla is small stream called Wadi El Tih bounded the New Cairo from south. These wadies are generally of dendritic and sub-parallel types.

Most of the surface of the area West G.C is covered with very gentle-dipping Tertiary-Quaternary strata including 6th of October City (Fig. 23 A). However, the area in some parts characterized by different topographic features mainly consists of Giza plateau (~120 m) and small folded and faulted Abu Roash complex and El Hassana Dom to the north of Giza pyramids (Fig. 10). The latter is highly manifested and controlled by both the lithologies and structures. The drainage lines have a very low degree of shaded relief (hill shading) (Fig. 4 and 11) on western side of the G.C and only Wadi Tafla south of the Pyramid could be detected.

Figure 10. Hassana Dom at Abu Roash area

 ISBN: 978-1-61208-251-6 

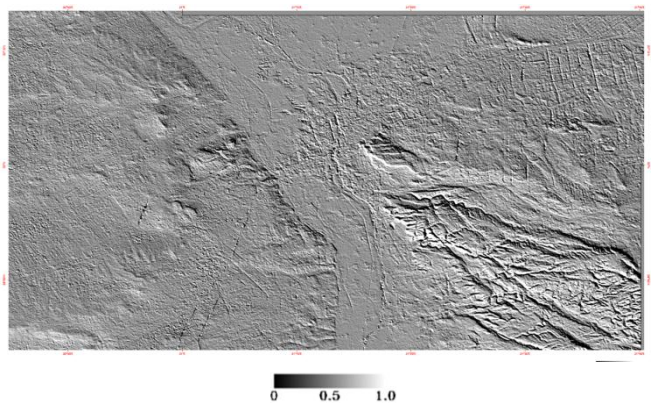



Figure 11. Shaded relief (hill shading) of the Greater Cairo area.

## V. CONCLUSIONS

Digital maps gain not only all ground and rock cover information but also are flexible for updating based on addition of new data and/or applications of advanced remote sensing techniques, which could be used for different purposes and to build a data base for large areas.

A study of Greater Cairo as the biggest city in MENA (Middle East North Africa) arid region shows that the geological setting might be one of the effective driving forces for further studies concerning the growing trends of the city and developing plans.

There are eleven rock types covering the Greater Cairo metropolis, composing mainly of carbonates and clastics with dispersed basaltic exposures. Although, these coverings are not dominated by any valued ores, some areas contain native structures and land forms. While the field observations and false composite high resolution satellite images (SPOT) showed degradation of protected areas; e.g. destroying land forms of El-Hassana Dom and misuse of Fores of Wood at Wadi Degla.

The statistical report of spectral signatures derived from PC merged TM image showed that the carbonate covers have higher numbers than clastics deposits (ex. Mean of band 1 addresses that more than 350 could be carbonate and between 350-320 could be clastics, while the basalt between 210-240 and wadi deposits reflects high value about 380).

Because the western part of G.C metropolis is dominated manly with clastics and characterized by a low relief and less roughness, the eastern part is built up mainly of carbonates and a rough relief, it can be mentioned that the western side has a higher opportunity for development than the eastern part.

## REFERENCES

[1] P. Ravi, “Remote Sensing Geology”, second edition, Springer 2003, pp. 1-3.

[2] A. Ibrahim, and M. Johari,“The Correlation of the Landsat TM Images Characteristic with the Geologic Information : A Preliminary Result of a Case Study in Langkawi Islands, Malaysia”, ACRS, 1997.

[3] M. Abd Manap , M. R. Firuz, W.N.Sulaima and N. Surip, “Application of Remote Sensing in the Identification of the Geological Terrain Features in Cameron Highlands, Malaysia “,Sains Malaysiana, vol. 39(1) , pp. 1–11, 2010.

[4] SS. Nalbant, and ö. Alpiekin, “The use of Landsat Thematic Mapper imagery for analysing lithology and structure of Korucu-Du la area in western Turkey”, International Journal of Remote Sensing Publication, vol. 16, issue 13, pp. 2357-2374, 1995.

[5] L. Fitzpatric, “-Producing Alaska Interim Land Cover Maps from Landsat Digital and Ancillary Data”, in Proceedings of the 11th Annual William T. Pecora Memorial Symposium: Satellite Land Remote Sensing: current programs and a look into the future American Society of Photogrammetry and Remote Sensing , pp. 339 – 347, 1987.

[6] M. Sultan, R. E. Arvidson and N. C. Sturchio, “Mapping of serpentinites in the Eastern Desert of Egypt by using Landsat thematic mapper data” The society of American Geology December, vol. 14, no. 12, pp. 995-999, 1986.

[7] S. Gad., T. Kusky, “ASTER spectral ratioing for lithological mapping in the Arabian-Nubian shield, the Neoproterozoic Wadi Kid area, Sinai, Egypt”,Gondwana research vol.11, pp. 326-335, 2007.

[8] S. l. Ingram, E. Gerg.., and H. Khald.., “Application of remote sensing for mapping surface geology in heavily vegetated cover in North Mississippi”, Journal of the Mississippi Academy of Sciences / Jan, 2003, Abstract.

[9] A. Osman, "Geotechnical and Structural studies on some new cities around Cairo” Cairo University, Faculty of Science, Geology Department. pp. 411-420, 2010.

[10] ERDAS Field Guide, 5th Edition, ERDAS Inc., Atlanta, Georgia, USA, 2010.

[11] K. V. Robert, “Fundamentals of geological and environmental remote sensing”, Prentice Hall series in geographic information science, University of California, 1997.

[12] J. R. Jensen,“Introductory digital image processing: a remote sensing perspective. Prentice-Hall, Upper Saddle River, NJ, pp. 318, 1996.

[13] J. A. Richards, “Remote sensing digital image analysis: an introduction “ (second edition).1993.

[14] I. A. M Frag, and M. M. Ismail,“Contribution to the stratigraphy of the Wadi Hof area (North-East of Helwan)”, Bull. Fac. Sci., Cairo, Uni., vol..27, pp. 221-239, 1959.

[15] R. Said, “The Geology of Egypt”, Elsevier Pub. Co., Amsterdam, 1962, p. 377.

[16] M. L. Abdelkhalek, M. A.. El Sharkawi, M. Darwish, ., M. Hagras, and A. Sehim, “Structural history of Abu Roash district, Western Desert, Egypt,” Journal of African Earth Science, vol 9, No 3/4, pp. 435-443, 1989.

[17] A. R. Moustafa, “Wernch tectonics in the north western desert of Egypt (Abu Roash area, West of Greater Cairo)”. Earth Sc. Ser. vol 2, 1988, pp. 1-16.

[18] A. R. Moustafa, F. El-Nahhas, and S. Abdeltawab, “Engineering geology of Mokattam city and vicinity, eastern Greater Cairo, Egypt.” Eng.Geol., vol. 31, pp. 327-344, 1991.

[19] A. R. Moustafa, A. Saoudi,I. M. Ibrahim,H. Molokhia, and B. Schwartz, B., “Structural setting and tectonic evolution of the Bahariya Depression, western Desert, Egypt”, Geo-Arabia, Bahrain, vol. 8, No. 1, pp. 91-124, 2003.

[20] A. H. Swedan, “A note on the geology of Greater Cairo area”, Annals of the geological survey of Egypt, vol XVII, pp 239-251, 1991.

[21] A. Sehim, “Cretaceous tectonics in Egypt”, Egypt. Jour. Geol., vol. 37, pp. 335-372, 1993.

[22] P. Norton, “Rock stratigraphic nomenclature of the Western Desert”, Pan American U.A.R Oil Co., Internal Report, p. 18, 1967,:

[23] N. M. Shukri, “Remarks on the geological structures of Egypt”, Bull. Soc. Geography.of Egypt, vol. 27, pp. 65-82, 1954.

 ISBN: 978-1-61208-251-6 

# Towards an ESB-Based Enterprise Integration Platform for Geospatial Web Services

Bruno Rienzi
InCo, Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay
brienzi@fing.edu.uy

Laura González
InCo, Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay
lauragon@fing.edu.uy

Raúl Ruggia
InCo, Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay
ruggia@fing.edu.uy

***Abstract*— Geographic Information Systems (GIS) have emerged to store, handle, analyze, and present geographic data to experts and casual users alike. As the number and scope of geo-enabled applications have increased considerably in recent years, new software solutions, like Internet Map Servers (IMS), have been incorporated into the enterprise software portfolio, giving birth to the "Enterprise GIS". Although these systems are based on standards for distributed geospatial processing and data exchange, through the use of Geospatial Web Services, they have to be complemented with other mechanisms to integrate them with business data and functionalities provided by traditional enterprise systems. In turn, an Enterprise Service Bus (ESB) is a standards-based integration platform, which provides mediation capabilities to address mismatches among applications regarding communication protocols and message formats, among others. This paper proposes an ESB-based reference platform which, leveraging its mediation capabilities, provides reusable geo-oriented integration mechanisms to deal with common challenges of integrating traditional enterprise systems with Geospatial Web Services.**

***Keywords-gis; enterprise applications; integration; geospatial web services; enterprise service bus.***

## I. INTRODUCTION

Traditionally, a Geographic Information System (GIS) has been defined as a system which integrates hardware, software, and data for capturing, managing, analyzing and displaying all forms of geographically referenced information (for a more extensive definition, see [1]). Looking back just a couple of decades, GIS was mainly provided by desktop applications, which stored data in files using proprietary formats. GIS was clearly geared towards experts. In the past decade, however, this situation has changed. It has been increasingly understood that location, as well as time, is a ubiquitous dimension of almost all data. As a result, many enterprise technologies, like relational databases and application servers have adopted GIS components to provide Geographic Information (GI) to a broader number of users. Enterprise GIS was born [2].

Like any other application within an organization, a GIS needs to be integrated with other software systems. In order to facilitate this integration, the Open Geospatial Consortium (OGC) has defined an architecture for distributed geospatial processing and data exchange, based on its own set of Web Services (WS) which are known as OWS (OGC Web Services).

In a simple scenario, an organization has its geographically-related data tied together in the same data source, typically a spatial database, and deploys off-the-shelf products, such as an Internet Map Server (IMS) that will query the spatial database to publish the GI, and a Map Viewer that will communicate with the IMS through OWS to allow user interaction. However, it is often the case that a map has to show additional information that comes from several data sources. For instance, a certain company could be interested in using an addresses map (or, more precisely, an addresses *layer*) that contains all the geo-coded addresses in a country and link those addresses to its clients data to perform certain analysis on the map. As the clients' data may spread over the company´s CRP databases and external systems (only accessible via WS), an IMS alone cannot cope with this integration requirement.

While several approaches [3] can be followed to integrate heterogeneous data and services in a complex scenario involving geographic and non-geographic information systems, the absence of standardized mechanisms leads to high costs and complexity in developing home-made integration solutions, which suffer from strong limitations in terms of service-oriented connectivity and reusability.

In this work, an Enterprise Service Bus (ESB) is used as the basis of an integration platform that addresses the aforementioned limitations in current systems. An ESB is a standards-based integration platform that combines messaging, WS, data transformation, and intelligent routing to reliably connect and coordinate the interaction of diverse applications [4]. An ESB provides a middle integration layer, with reusable integration and communication logic, which helps to address mismatches among applications regarding communication protocols, message formats, and quality of service (QoS), among others [5].

This paper proposes an ESB-based platform which extends the basic mediation capabilities with reusable geo-oriented integration mechanisms to deal with common challenges of integrating traditional enterprise systems with OWS, focusing on providing a reference architecture for this platform and describing various concrete geo-oriented integration mechanisms.

The rest of the paper is organized as follows. Section II provides background concepts. Section III presents a solution approach to address the aforementioned integration

 ISBN: 978-1-61208-251-6

challenges. Section IV proposes and describes various concrete geo-oriented integration mechanisms. Section V presents implementation details. Finally, Section VI presents conclusions and future work.

## II. BACKGROUND

This section provides background on OWS and ESB mediation patterns. These technologies constitute the basis to build the proposed solution.

### A. OGC Geospatial Web Services

OWS [6] have coarse-grained interfaces with a few stateless operations. They use HTTP and XML, but not SOAP or WSDL (the W3C standards for WS).

This paper focuses on two OWS standards: Web Map Service (WMS) and Web Feature Service (WFS). WMS [7] produces maps dynamically. A map, in WMS terms, is a portrayal of GI as a digital image file, which is the result of overlaying several geographic layers. A layer is collection of features of the same type (e.g., a roads layer). A feature [8] is an abstraction of a real world phenomenon (e.g., a road). WMS provides two specific operations: the mandatory GetMap and the optional GetFeatureInfo. A GetMap request specifies, among others, the layers to make up the map, the coordinate reference system (e.g., Universal Transverse Mercator) and the geographic area (e.g., the rectangular boundaries of a country). A GetMap response is the map itself. On the other side, a GetFeatureInfo request specifies the coordinates of a point and a GetFeatureInfo response brings the attribute data of the features that contain or are near that point (as long as the features belong to a *queryable* layer). The attribute data of a road could be its name, type, average traffic, etc.

WFS [9] offers direct fine-grained access to GI at the feature and feature property (attribute data) levels. It allows clients to retrieve, create, modify and delete features, using XML-based messages, independently of the storage. This paper concentrates on two WFS mandatory operations: DescribeFeatureType and GetFeature. DescribeFeatureType returns an XML schema defining the feature type (i.e., the names and data types of the attributes that define the structure of the features in a given layer). GetFeature returns the instances of a certain feature type that match a geographical filter (e.g., return all roads within a geographic area). Both WMS and WFS have a GetCapabilities operation to return the service metadata (e.g., the layers list, supported response formats, supported versions, the URLs to invoke the other operations, etc.).

Internet Map Servers are server-side applications that implement and expose WMS and WFS services, among others.

### B. ESB Mediation Patterns

Within an ESB-based platform, services and applications communicate by sending messages through the ESB. Messages are processed by mediations flows which can apply to them different mediation operations (e.g., routing). In this way, the ESB can ensure that applications and services connect successfully [10][11]. Although mediations are not formally restricted in what they can do, there are a set of basic patterns, known as mediation patterns, that are seen repeatedly and have been documented [4][10][12][13].

Transformation patterns deal with the runtime transformation of messages. Routing patterns dynamically determines the message path according to different factors. For example, the Content-Based Routing (CBR) determines the message path based on its content. The Splitter Pattern breaks out a message into a series of individual messages. The Aggregator Pattern receives multiple messages and when a given set of messages is complete, a single message is returned consolidating their content. Finally, the Cache Pattern returns messages which were previously stored and returned as a response for the given request [13][14].

Figure 1 presents a summary of these mediation patterns and a graphical representation for them, introduced in [5].

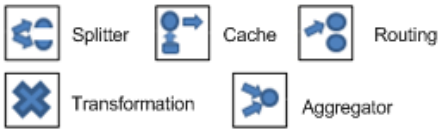



Figure 1. Summary of Mediation Patterns.

## III. SOLUTION APPROACH

The proposed solution consists of an ESB-based Enterprise Integration Platform (EEIP) which provides reusable geo-oriented integration mechanisms to deal with common challenges of integrating OWS with enterprise applications. Figure 2 presents the general architecture of the platform and the external systems with which it interacts. The Basic and Complex Geo-oriented Integration Mechanisms are proposed and specified in this paper, while the underlying mediation mechanisms are usually available in ESB solutions. The platform acts as a broker between clients and servers (either OWS servers or enterprise systems). Clients send their requests to the ESB and the ESB routes them to real servers, applying some mediation flows, in a transparent fashion. These flows return value-added responses that could not be obtained by directly consulting the servers.

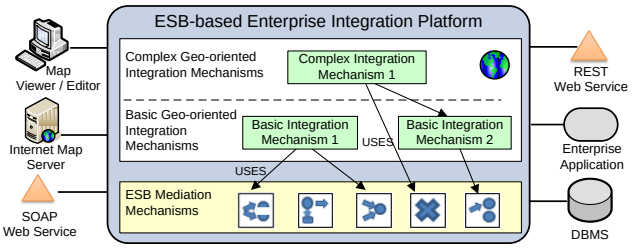



Figure 2. ESB-based Enterprise Integration Platform for Geospatial Web Services.

 ISBN: 978-1-61208-251-6

The ESB Mediation Mechanisms correspond to the mediation patterns described in section II.B (i.e., routing, splitter, etc.). These mechanisms are usually included in ESB products and provide reusable solutions to deal with general integration and communication requirements.

The Basic Geo-oriented Integration Mechanisms are higher-level mechanisms, built on top of the previous ones, which provide either geo-oriented utilities or reusable solutions to integrate OWS with enterprise systems.

The Complex Geo-oriented Integration Mechanisms are mediation mechanisms, built on top of the two previous ones, which provide higher level reusable solutions to integrate OWS with enterprise applications.

Additionally, the platform interacts with various external systems including IMSs, enterprise applications, SOAP / REST WS and map viewers / editors, among others. These systems provide information or consume services exposed in the platform. For instance, the platform can consume a SOAP WS to obtain business data from an enterprise system or a map viewer can consume a WMS interface exposed as a service in the platform.

## IV. GEO-ORIENTED INTEGRATION MECHANISMS

This section presents and describes two complex geo-oriented integration mechanisms: WMS Enricher and SOAP-WMS Wrapper. Each mechanism is first described through a general description which includes: a motivation (i.e., why the mechanism is needed), a set of constrains (i.e., restrictions to be considered for the solution), a high level solution and an application example.

An in-detail solution is also provided for each mechanism. This in-detail solution is specified through mediation flows built on top of the ESB mediation mechanisms and some basic geo-oriented integration mechanisms, which are also described.

Finally, a set of variants (i.e., modifications or improvements) is analyzed for each mechanism.

### A. WMS Enricher

The WMS Enricher is a complex geo-oriented integration mechanism which addresses the issues of integrating heterogeneous systems in GIS-based applications. To this end, it includes mediation flows to complement WMS responses with business data which are absent in the IMS.

#### 1) General Description

TABLE I. presents the general description for the WMS Enricher mechanism.

TABLE I. WMS ENRICHER GENERAL DESCRIPTION

| **WMS Enricher** | |
|---|---|
| **Motivation** | An organisation has an IMS that is accessed by external WMS Clients. This organization wishes to complement its GI (accessible through the IMS) with related business data located in an enterprise system (accesible through a WS). |
| **Constraints** | Since the WMS Clients are external and the organization cannot modify them, the mechanism implementation must be transparent to those clients. |
| **Solution** | A mediation flow which enriches WMS responses is implemented within the EEIP. This flow is responsible for handling WMS requests, splitting them into business-dependant requests, forwarding these requests to the appropiate destination (IMS o enterprise system), aggregate the business-dependant responses and return the enriched WMS response to the client. |
| **Application Example** | The Ministry of Tourism has an IMS which publishes a Points of Interest (POI) layer that it wants to complement with the number of visitors in the last month of each POI. This information is provided by an independent company through a SOAP WS. |

#### 2) In-detail Solution

Figure 3 presents a high level view of the WMS Enricher internals, which leverage a set of basic reusable geo-oriented mechanisms.

Instead of directly interacting with an IMS, a WMS client sends WMS requests to the EEIP. These requests are processed differently according to the operation involved (GetCapabilities, GetMap, GetFeatureInfo).

When the EEIP receives a WMS request, it creates an ESB message containing the request.

This message is first processed by the GeoEntryPoint mechanism (1), which obtains the invoked operation from the WMS request, set a message property with this operation and returns the modified message.

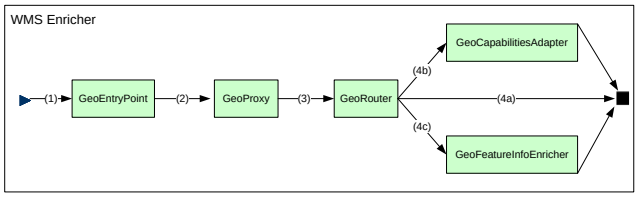



Figure 3. Aggregating data from WMS and other external systems.

Then, the message is processed by the GeoProxy mechanism (2), which works as a gateway to the backing IMS. It modifies the received message replacing the request with the response returned by the IMS.

Next, the message is processed by the GeoRouter mechanism (3). This mechanism routes the message according to the name of the invoked operation, which was previously stored in a message property.

If the operation is GetMap, the GeoRouter return the response to the invoking client (4a). If the operation is GetCapabilities, the GeoRouter routes the message to GeoCapabilititesAdapter (4b). This mechanism is needed because the original GetCapabilities response contains the URLs to invoke the other operations (GetMap and GetFeatureInfo), and as a result, these have to be replaced so that clients can invoke them through the EEIP in the subsequent requests. If the operation is GetFeatureInfo, the response is enriched using other sources, for example, SOAP and REST WS. To this end, the GeoRouter routes the request to the GeoFeatureInfoEnricher mechanism (4c), which receives an ESB message containing a WMS response and, more precisely, the IDs of the features involved. The GeoFeatureInfoEnricher has the purpose of invoking the other external sources, using those IDs, to obtain additional

 ISBN: 978-1-61208-251-6

information of the features. It has to aggregate the responses obtained from these sources with the original IMS response.

In the following paragraphs, the basic geo-oriented integration mechanisms that have been mentioned are described in the depth.

**GeoEntryPoint** is a basic geo-oriented integration mechanism which receives ESB messages containing an IMS request, obtains data from this request and stores them in message properties for later processing. In order to use this mechanism, it is necessary to specify what information has to be obtained from the requests (e.g., invoked operation).

**GeoProxy** is a basic geo-oriented integration mechanism which has the purpose of invoking an OWS operation on an IMS. It receives a message containing a request and returns a modified message replacing the request with the response obtained from the IMS. In summary, this mechanism acts as an HTTP gateway. In order to use this mechanism, the URL of the IMS has to be specified.

**GeoRouter** is a basic geo-oriented integration mechanism which uses the ESB Routing, more precisely, the CBR. It routes messages to the appropriate destination according to message properties. To completely specify the behavior of this mechanism, a set of (property, value, destination) triples has to be specified.

**GeoCapabilitiesAdapter** is a basic geo-oriented integration mechanism which uses the ESB Transformation in order to replace all the original URLs published in the GetCapabilities response with the equivalent URLs that handle the requests through the EEIP. For instance, Figure 4 shows a fragment of a GetCapabilities response, in which the OnlineResource element indicates the URL where the GetFeatureInfo operation must be invoked using the HTTP Get method. Since this URL points to the IMS, GeoCapabilitiesAdapter must replace it with the corresponding URL in the EEIP (see Figure 5). To completely specify the behavior of this mechanism the required parameter is the list of pairs (original URL, adapted URL).

```
<GetFeatureInfo>
    <Format>text/plain</Format>
    <Format>application/vnd.ogc.gml</Format>
    <Format>text/html</Format>
    <DCPType>
        <HTTP>
            <Get>
                <OnlineResource xlink:type="simple"
                    xlink:href="http://imshost:8080/ims/ows?SERVICE=WMS&" />
            </Get>
        </HTTP>
    </DCPType>
</GetFeatureInfo>
```

Figure 4. GetCapabilities response fragment showing the original URL of the GetFeatureInfo operation.

```
<GetFeatureInfo>
    <Format>text/plain</Format>
    <Format>application/vnd.ogc.gml</Format>
    <Format>text/html</Format>
    <DCPType>
        <HTTP>
            <Get>
                <OnlineResource xlink:type="simple"
                    xlink:href="http://eeiphost:8080/eeip/ows?SERVICE=WMS&" />
            </Get>
        </HTTP>
    </DCPType>
</GetFeatureInfo>
```

Figure 5. GetCapabilities response fragment showing the adapted URL of the GetFeatureInfo operation.

**GeoFeatureInfoEnricher** is a basic geo-oriented integration mechanism which uses the ESB Splitter, Aggregator and Transformation in order to obtain business data from enterprise applications and consolidate the responses into a unique GetFeatureInfo response. Figure 6 shows the mediations flow that occurs inside this mechanism. As in [5], YAWL [15] is used to uniformly represent this mediation flow among ESB products.

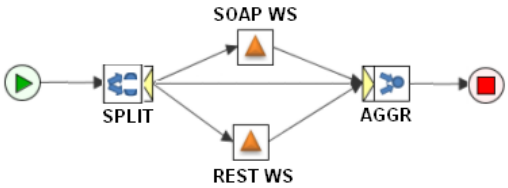



Figure 6. GeoFeatureEnricher Mechanism as a YAWL Net.

In the first place, a Splitter receives a message containing the IMS response and sends the requests to the external sources (e.g., a SOAP WS). To accomplish this task, some parameters have to be specified for each external source: its type (e.g., SOAP WS, REST WS, EJB, etc.), its address (e.g., an URL), the operation to invoke and the attribute name where the feature ID has to be set to invoke the operation. The Splitter has the knowledge to send the required information for each external source type. Also, the WS SOAP WS and REST WS mechanisms know how to build requests for each type of WS, respectively.

When all the responses are received, an Aggregator is in charge of consolidating them into a single GetFeatureInfo response. To accomplish this, some parameters have to be specified for each layer: the layer name, a set of triples of the form (attribute, source, locator) where *attribute* is the name of an attribute to be added to the response, *source* is an external source from where the attribute value is obtained, and *locator* is source-type dependant way to get the value from the external source response (i.e., an XPath or XQuery expression in the case of a SOAP response, etc.).

*3) Variants*

Following the same approach that has been applied for the WMS Enricher, a possible variant for this mechanism is to derivate a similar solution to enrich a non-transactional WFS, i.e., the WFS Enricher. In this scenario, a WFS client interacts with the EEIP issuing GetCapabilities, DescribeFeatureType and GetFeature requests. In the case of the WFS GetCapabilities operation, the same adaptation that is performed in the WMS Enricher applies without modification. In the case of DescribeFeatureType, the XML schema that is returned has to be augmented with the attributes that are not part of the original feature type. The GeoFeatureTypeAdapter mechanism is defined to accomplish this task. In the case of GetFeature, a similar mechanism to the GeoFeatureInfoEnricher is defined, the GeoFeatureEnricher, which queries the external applications to retrieve the additional data for each feature in the

 ISBN: 978-1-61208-251-6 

GetFeature response. Figure 7 presents a high level view of this variant.

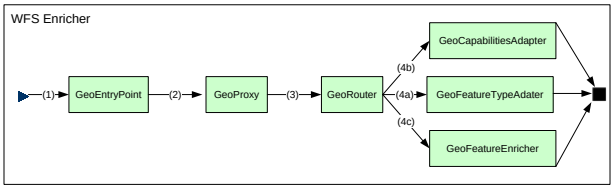



Figure 7. Aggregating data from a basic WFS and other external systems.

Given that this kind of integration solution has a clear impact on performance, another variant for the WMS and WFS Enricher mechanisms is to leverage previous processing to return a response. A possible strategy to do that is using previously returned information, through a cache mechanism [16]. In this case, the GeoEntryPoint could query the Cache before sending a request to the GeoProxy. If the response is found in the cache, it is returned to the client instead of following the usual flow.

### *B. SOAP-WMS Wrapper*

The SOAP-WMS Wrapper is a complex geo-oriented integration mechanism which serves the purpose of publishing WMS services using W3C Web Services Standards (SOAP and WSDL). In this way, they can be integrated into general-purpose business-to-business (B2B) processes that rely on those standards and leverage an extensive stack of related specifications to address advanced concerns such as security, reliability, discoverability, orchestration, etc.(see discussion on [17]).

#### *1) General Description*

TABLE II. presents the general description for the SOAP-WMS Wrapper mechanism.

TABLE II. SOAP-WMS WRAPPER GENERAL DESCRIPTION

| **SOAP-WMS Wrapper** | |
|---|---|
| **Motivation** | An organisation wishes to incorporate a IMS into its existing Web Services infrastructure, so as to make GI accesible to its existing business processes and applications. |
| **Constraints** | Since the organisation´s processes and applications are already implemented using SOAP WS and related WS-* specifications to meet some advanced constraints (e.g., WS-Security, WS-BPEL, etc.), it is not possible or desirable to consume WMS services using plain HTTP messages interchange. |
| **Solution** | A mediation flow which wraps WMS using SOAP WS is implemented within the EEIP. This flow is responsible for handling SOAP-WMS requests, translating those requests into standard WMS requests, forwarding these requests to the IMS, translating the WMS responses into SOAP-WMS and returning them to the client. |
| **Application Example** | The Ministry of Energy and Mineral Resources has implemented a BPMS, based on SOAP WS, to manage its business processes, which include awarding grants for mine prospection, exploration and exploitation. In a first effort to incorporate GI into the system, an activity has been defined in an early stage of the process in which an analyst can visualize the mine area affected by the grant request and the protected areas layer (where mining is prohibited) on the same map. |

#### *2) In-detail Solution*

The SOAP version of WMS is published in the EEIP following the guidelines from [18]. Figure 8 shows a high level view of the SOAP-WMS Wrapper.

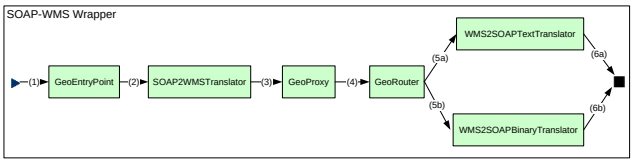



Figure 8. Wrapping WMS into SOAP WS.

Once a SOAP request arrives at the EEIP an ESB message is created containing this SOAP request. The message is first processed by the GeoEntryPoint mechanism (1) which obtains, from the request, the operation that is being invoked and stores its name in a message property. The modified message is then processed by the SOAP2WMSTranslator (2) which has to decode the SOAP request, create a standard WMS request, either using the KVP or XML encoding, and replace, in the message, the SOAP request with a standard WMS request. The output message is then processed by the GeoProxy (3) which invokes the operation on the IMS, receives the result and replaces the request with the response. The ESB message is then processed by the GeoRouter (4) according to the operation that was invoked. If the operation returns a text response (i.e., GetCapabilities and GetFeatureInfo), it is sent to the WMS2SOAPTextTranslator (5a), if it returns a binary response (i.e., GetMap) it is sent to the WMS2SOAPBinaryTranslator (5b). These two mechanisms work in the same way, by decoding the response and creating a SOAP message to return to the client (6a and 6b).

**SOAP2WMSTranslator** is a basic geo-oriented integration mechanism which uses the ESB Transformation in order to convert a SOAP message into a WMS request. This mechanism can be configured to either generate a KVP o XML encoded requests, to be used by the HTTP Get or HTTP Post methods respectively [19]. The Transformation mechanism is dependent on the ESB platform of choice, but typically includes XML-based XSLT transformations or object marshalling/unmarshalling, the latter being programming language bound. To completely specify the behavior of this mechanism the required parameters are the type of encoding (KVP, XML) and the URLs of the IMS that handle each type of request.

**WMS2SOAPTextTranslator** is a basic geo-oriented integration mechanism which uses the ESB Transformation in order to convert a WMS response into a SOAP message. Only text-format responses are processed by this mechanism, i.e., the responses of the GetCapabilities and GetFeatureInfo operations.

 ISBN: 978-1-61208-251-6

**WMS2SOAPBinaryTranslator** is a basic geo-oriented integration mechanism which uses the ESB Transformation in order to convert a WMS response into a SOAP message. Only binary-format responses are processed by this mechanism, i.e., the responses of the GetMap operation. This mechanism uses the Message Transmission Optimization Mechanism (MTOM) to attach the binary data to the SOAP message.

*3) Variants*

An analogue mechanism to the SOAP-WMS Wrapper, could make possible to expose other OWS´s as SOAP WS. In the case of WFS, its inclusion in a Service Oriented Architecture (SOA) using the EEIP could go even further than just visualizing GI, as is the case with WMS, since WFS offers a complete interface to perform advanced queries and transactions on GI.

## V. IMPLEMENTATION DETAILS

In order to show the feasibility of the proposed approach, some prototypes have been developed based on the JBoss ESB product, using GeoServer as the IMS. These prototypes have also allowed identifying and analyzing key implementation aspects.

In particular, the WMS Enricher mechanism was developed by leveraging various built-in features of JBoss ESB, like its CBR and aggregator features [20]. The variants of the WMS Enricher mechanism (WFS Enricher and Cache) were also implemented [21] leveraging these features. However, some extra work was required to implement a cache mechanism, given that JBoss ESB does not natively provide this feature.

Finally, the WMS SOAP Wrapper was also successfully implemented with JBoss ESB [22] leveraging its native features, like a SOAP Processor to perform the marshalling and unmarshalling.

## VI. CONCLUSION AND FUTURE WORK

This paper addressed the issues of integrating GIS with enterprise systems to build large-scale Information Systems that use GI in a broader business context. More concretely, it presents an ESB-based Integration Platform which provides generic Geo-oriented Integration Mechanisms to facilitate the integration of GIS, and more specifically OWS, with traditional enterprise counterparts. Those mechanisms are described and specified in detail through mediation flows, which extend mediation patterns commonly supported in ESB products (e.g., Splitter and Aggregator).

Also, the development of various prototypes has shown the feasibility of the proposed approach and mechanisms.

The main contributions of this paper consists of the specification and implementation of a reference platform and concrete mechanisms, based on a general purpose ESB, that address key aspects of integrating heterogeneous systems in a GIS environment. The specification and implementation of this kind of platform has not been tackled by previous works to the best of authors' knowledge.

As part of an ongoing project, this work aims to be a step forward in developing a comprehensive platform which facilitates organizations the task of integrating GIS with their traditional enterprise systems.

Future work consists in improvements to these mechanisms, the design and development of new ones (e.g., a SOAP-WFS Wrapper, a Transactional WFS Enricher, etc.), and the exploration of approaches to carry out the integration with an ESB-based e-Government platform.

## REFERENCES


[1] D. J. Maguire, "An overview and definition of GIS" in D. J. Maguire, M. F. Goodchild, & D. W. Rhind, Eds., Geographical Information Systems: Principles and Applications (Vol. 1), Harlow, UK: Longman Group, pp. 9-20, 1991

[2] M. S. Witkowski, P. M. Rich, G. N. Keating, and L. A. N. Laboratory, "A prototype for enterprise GIS", Los Alamos National Laboratory, 2004, electronic resource.

[3] B. Rienzi, R. Sosa, P. Foti, and L. González, "Benefits and challenges of using geographic information systems to enhance social security services.," presented at the 6th International Policy and Research Conference on Social Security, Luxembourg, 2010.

[4] D. Chappell, "Enterprise Service Bus: Theory in Practice", O'Reilly Media, 2004.

[5] L. González and R. Ruggia, "Addressing QoS issues in service based systems through an adaptive ESB infrastructure", Proceedings of the 6th Workshop on Middleware for Service Oriented Computing - MW4SOC '11. pp. 1–7, Lisbon, Portugal, 2011.

[6] A. Whiteside, Ed., "OpenGIS Web Services Architecture Description", Open GIS Consortium Inc., 2005.

[7] J. de la Beaujardiere, Ed., "OpenGIS Web Map Server Implementation Specification", Open GIS Consortium Inc., 2006.

[8] C. Kottman, C. Reed, Eds., "The OpenGIS Abstract Specification. Topic 5: Features", Open GIS Consortium Inc., 2009.

[9] P. A. Vretanos, Ed., " OpenGIS Web Feature Service 2.0 Interface Standard", Open GIS Consortium Inc., 2010.

[10] M. T. Schmidt, B. Hutchison, P. Lambros, R. Phippen, "The enterprise service bus: making service-oriented architecture real", IBM Syst. J. 44, 781–797, 2005.

[11] M. Papazoglou, "Web Services: Principles and Technology", Prentice Hall, 2007.

[12] H. Wylie, P. Lambros, "Enterprise Connectivity Patterns: Implementing integration solutions with IBM's Enterprise Service Bus products", http://www.ibm.com/developerworks/library/ws-enterpriseconnectivitypatterns/index.html.

[13] G. Hohpe and B. Woolf, "Enterprise Integration Patterns: Designing, Building, and De-ploying Messaging Solutions", Addison-Wesley Professional, 2003.

[14] R. Yan Fang, F. Ru, T. Zhong, L. Eoin, S. Harini, T. Banks and L. He, "Cache mediation pattern specification: an overview", http://www.ibm.com/developerworks/webservices/library/ws-soa-cachemed/.

[15] A.H.M. ter Hofstede, W.M.P. van der Aalst, M. Adams and N. Russell: Modern Business Process Automation: YAWL and its Support Environment. Springer, 2009.

[16] L. González and R. Ruggia, "Towards dynamic adaptation within an ESB-based service infrastructure layer", Proceedings of the 3rd International Workshop on Monitoring, Adaptation and Beyond. pp. 40–47. ACM, New York, NY, USA, 2010.


 ISBN: 978-1-61208-251-6 

[17] J. T. Sample, K. Shaw, S. Tu, and M. Abdelguerfi, Eds., “Geospatial Web Services: Bridging the Gap Between OGC and Web Services,” in Geospatial Services and Applications for the Internet, Springer US, 2008, pp. 72–92.

[18] J. Sonnet and C. Savage, “OWS 1.2 SOAP Experiment Report”, Open GIS Consortium Inc., 2003.

[19] R. Gartmann and B. Schäffer, “OpenGIS Wrapping OGC HTTP-GET and -POST Services with SOAP - Discussion Paper”, Open GIS Consortium Inc., 2008.

[20] E. Rosa, J. Davison and G. Schnyder, “Integración de Datos Empresariales en Servicios Geográficos a través de un ESB”, TSIG, 2010, unpublished,
http://www.fing.edu.uy/inco/cursos/tsi/TSIG/papers/wms-enr.pdf [accessed: 2012-12-21]

[21] E. Penna, V. Perez and M. Steffen, “WFS en ESB con Fuente Externa y Cache”, 2011, unpublished,
http://www.fing.edu.uy/inco/cursos/tsi/TSIG/papers/wfs-enr.pdf [accessed: 2012-12-21]

[22] P. García, J. Laborde and M. Rodríguez, “WMS SOAP con ESB”, TSIG, 2011, unpublished,
http://www.fing.edu.uy/inco/cursos/tsi/TSIG/papers/wms-soap.pdf [accessed: 2012-12-21]

 ISBN: 978-1-61208-251-6 

# A GIS-based Spatial Analysis of Health care Facilities in Yola, Nigeria

Abdurrahman Belel Ismaila
Department of Geodetic & Geographic Information Technologies
Middle East Technical University
Ankara, Turkey
belelismaila@yahoo.com

Nurünnisa Usul
Department of Civil Engineering
Middle East Technical University
Ankara, Turkey
nurusul@metu.edu.tr

***Abstract*—In recent years, health care provision policies in Nigeria have addressed distribution and spatial equity questions at a gross or regional level, but have neglected to address the distribution of health care facilities within cities. This paper explores the potential use of GIS for modeling the spatial distribution and accessibility of the health care delivery system in Yola. Several digital and non-digital data sets were collected and transformed into GIS data. Spatial analysis tools, including symbols, overlay operations, Kernel Density Estimations (KDE), buffer operations, and a raster calculator were used for the analysis. All identified public and private facilities were classified as primary, secondary, or tertiary. The majority of these facilities were concentrated in Jimeta. The study also produced the three following accessibility models: (i) the distance to the health facility, (ii) the health facility-to-population ratios, and (iii) the physician-to-population ratios. Based on this analysis, it was concluded that a gross inadequacy exists in terms of health care facilities and physicians. Thus, these results identify the need for urgent improvements in the Yola health care delivery system, including the construction of new facilities, upgrades for existing facilities, increased physician employment, and the adoption of GIS technology by Yola health care planners and policy makers for effective planning and resource allocation.**

***Keywords-GIS; Spatial analysis; Health care facilities; Yola; Nigeria.***

## I. INTRODUCTION

The idea that all people are entitled to have their physical needs satisfied is at the heart of the human rights movement. This includes the right to survive and to live without preventable suffering. The Declaration of Human Rights details the right to adequate health in Article 25 as follows:

> "Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care…" [1].

Nigeria recognises the right to health and has committed itself to health protection by assuming obligations from international treaties and domestic legislations that mandate specific conduct regarding the health of individuals within Nigeria. Prior to the economic efforts in the mid-1980s, the health sector witnessed robust growth, mainly because of unfettered government support and assistance from international donor agencies. During this time frame, access to free health care was readily available at public hospitals and clinics in urban areas. However, this positive development suddenly stopped by 1985 because of multiple factors. Of these factors, the two most notable factors were the precipitous economic decline and the military usurpation of power. The military usurpation of power marked the genesis of many intractable challenges that impacted the Nigerian health system [2]. Since then, the geographic mal-distribution of regional, urban, and rural health care facilities in Nigeria has occurred. Moreover, the existing facilities are haphazardly distributed. This situation may be corrected by using reliable data with the Geographic Information System (GIS). These systems contain important tools that can help in health delivery service planning and decision-making processes. These tools include database management, planning, risk assessment, service area mapping, location identification, and accessibility tools [3]. For example, [4, 5] conducted hospital service area delineations, [6] calculated changes in the person-to-bed ratios within hospital catchment areas, [7] proposed a new location-allocation structure for primary health care centres by using a multi-criteria approach in a GIS environment, and [8, 9] computed the accessibility of physician locations and the physician-to-population ratio. The same principle was applied by [10], except non-spatial factors of health care accessibility were added, such as age and social class. [11, 12] used gravity models to define the population flow to health centres and to produce a combined accessibility indicator of distance and availability. Several of these studies are well acknowledged by researchers from Western European and other developed countries. In contrast, most health authorities, practitioners and researchers in Nigeria have not explored the potential use of GIS for improving the performance of essential public health services.

Thus, the aim of this study was to explore the use of GIS for modelling the spatial distribution and accessibility of the health care delivery system in Yola, Nigeria. This study was conducted to identify the underserved and over served areas and to create a model to guide health care planners and regulators in their creation of future proposals. The specific objectives of the study are to *i)* inventory all public and private health facilities, *ii)* create a GIS database of the existing health care centres, *iii)* characterise and identify the spatial pattern of the existing health care facilities, *iv)* map

 ISBN: 978-1-61208-251-6 

the health facility services, and *v)* model the accessibility of health care facilities for Yola residents.

Considering the advantages of GIS relative to traditional analysis methods, the major contribution of this study is to demonstrate that spatial GIS information can be used to reach desired quality decisions in a short time frame and at a low cost. In addition, GIS will enable the establishment of a health care facility database, which can be easily retrieved and analysed at any time.

## II. MATERIALS AND METHODS

### A. Study Area

Yola is the administrative capital of Adamawa State, Nigeria. Yola is a twin settlement that encompasses Jimeta (the administrative and commercial centre) and Yola-Town (the traditional settlement). Yola is located at a latitude of 9°14″ N and a longitude of 12°28′ E (Figure 1), and has a total population of 395,871. Specifically, Jimeta and Yola-Town have populations of 199,674 and 196,197, respectively [13]. The estimated population projection for 2011 is 399,598.

Yola has a tropical climate with rainy and dry seasons. The maximum and minimum temperatures can reach 40°C and 18°C in April and between December and January, respectively. The mean annual rainfall is less than 1,000 mm [14].

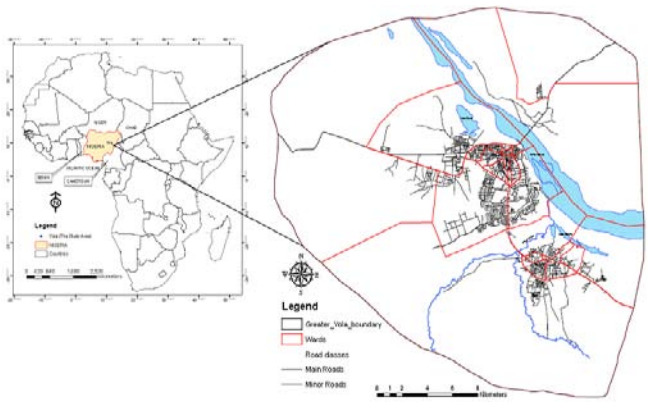



Figure 1. The study area.

### B. Data and Methods

Several digital and non-digital data sets that contain spatial data were collected through a literature review and from field work. The spatial data sets included analogue maps of the road network and administrative boundaries, global positioning system (GPS) coordinates of strategic road junctions (Table I) [15], and a Google earth satellite image of the study area. The GPS coordinates were used as a georeference for the analogue maps. The Google image was used to update the analogue maps and to identify the health facility locations. To depict the service supply environment, an inventory of all health care facilities (public and private) in the study area was conducted. This inventory was conducted by compiling an exhaustive list of facilities that were documented by [16] and the Adamawa State Ministry Health and Health Service Management Board in their official records.

TABLE I. THE GPS COORDINATES OF THE YOLA ROAD JUNCTIONS [15].

| Road Junctions | Northing | Easting | Altitude (meters) |
|---|---|---|---|
| Abuja Road/Modibbo Adama Way | 1019260 | 221732 | 123.72 |
| Abuja Road/Polo Road Junction | 1019062 | 223734 | 129.30 |
| Abuja Road/Modibbo Adama Way | 1018700 | 224418 | 134.42 |
| Modibbo Adama way/Bye pass Junction | 1017959 | 225980 | 143.72 |
| Mafias Quarters Road/Bye – pass | 1017551 | 224029 | 144.42 |
| Madaki Street/Bye – pass Junction | 1017397 | 223757 | 153.45 |
| Madaki Street/Modibbo Adama Way | 1018452 | 223818 | 153.02 |
| Sokoto Street/Modibbo Adama way | 1018549 | 223374 | 145.58 |
| Chiroma street/Modibbo Adama way | 1018473 | 222953 | 148.60 |
| Modibbo Raji/Bye – pass (Phase I Road Junction) | 1017483 | 223298 | 141.63 |
| Phase II Road/Bye – pass | 1017513 | 227780 | 146.74 |
| Abdullahi Bashir/Yola Road (80 Unit Junction) | 1021929 | 220437 | 136.51 |
| Police Round About | 1024445 | 220536 | 157.21 |
| Justice Buba Ardo/Ibrahim Kashim (Gov't House Junction) | 1024520 | 221890 | 169.30 |
| Zaki Crescent/Abdullahi Bashir Road | 1023508 | 222318 | 160.93 |

This study used the ArcGIS® 9.3 software. Thus, all data sets were transformed into the ArcGIS® data format (shapefile). Each of the analogue maps was scanned and georeferenced to UTM zone 32 N and to the "Minna-Nigeria" datum by using the GPS data (Table 1) as ground control points. In addition, the spatial data were captured and organised into different layers with on-screen digitisation as described below. Figure 2 shows the methodology schematic. Finally, a database was created and filled for each of the layers.

*i.* The road network layer: describes the road network patterns and contains the road name, length, width, class, and condition. This layer enables health planners to understand the spread of a city network and can be overlaid onto a health centre location map to examine their relationships.

*ii.* The health care facility location layer: contains a record of health centre locations in Yola. In addition, this layer contains the following attributes: (a) general information regarding the health facility (including name, ownership, ward and local government, legal status, and bed capacity), and (b) medical personnel data (for physicians, pharmacists, technicians, and nurses/midwives). Health centres can be classified by any of these attributes and then used for comparing the functional differences between the different health centres.

*iii.* The administrative boundary or service extent of the health centre layer: is a polygon layer that divides Yola into 20 different administrative wards. Each ward shows the spatial extent of the health

 ISBN: 978-1-61208-251-6 

centres. Attributes of population, area, and density were linked to this layer. Because population census data on a ward basis is lacking, the 2007 voters register from the Independent Nigerian Electoral Commission (INEC) [17] and the 2006 population census data of Yola were used to derive the ward populations. Population data help health planners to define the catchment area of each health centre and indicate whether these centres can offer adequately serve a population of a certain size.

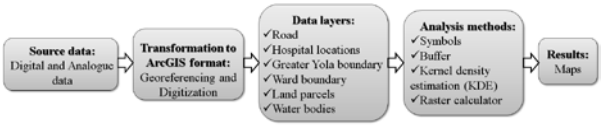



Figure 2. Methodology.

## C. Methods

*i.* *Symbols:* are used on maps to represent various geographic phenomena that are related to location, distance, volume, movement, function, process, and correlation. These phenomena can be classified into the four following basic categories: *point* (non-dimensional data), *line* (one-dimensional data), *area* (two-dimensional data), and *volume* (three-dimensional data) [18]. In this study, the *point* category is used to represent the health care facilities.

*ii.* *Overlay operation:* is a spatial operation that combines different geographic layers to generate new information. This operation is performed in raster and vector domains by using arithmetic, boolean, and relational operators [19]. However, in this study, this operation is only applied in the vector domain. Detailed descriptions of the overlay operations can be found in standard GIS textbooks, such as those by [20, 21, 22]. Furthermore, [23, 24] describe these principles in more detail and [25] detail some of the key overlay algorithms.

*iii.* *Kernel density estimation (KDE):* estimates the density by using a "kernel function" *K* and a smoothing parameter *h.* Generally, the estimated density does not depend on the kernel function choice, but is strongly affected by the smoothing parameter value. Larger smoothing parameters obscure small irregularities and smaller smoothing parameters result in a bumpy density surface [26]. There are two types of KDE functions, fixed and adaptive kernels. The fixed kernel function is usually less computationally intensive than the adaptive kernel function and uses an optimal spatial kernel (bandwidth) over the study space. Unfortunately, this method produces high local estimation variance in areas where data are sparse, which may mask subtle local variations in areas where data are dense [27, 28, 29]. In contrast, the adaptive kernel function ensures that a certain number of nearest neighbours are considered as local samples, which better represents the degree of spatial heterogeneity [28]. In this study, the Gaussian adaptive kernel function found in the ArcGIS® 9.3 Spatial Analyst Tool was used. The general form of the KDE is given by [30] as follows:

$$\lambda(s) = \sum_{i=1}^{n} \frac{1}{\pi r^2} k\left(\frac{d_{is}}{r}\right) \quad (1)$$

where $\lambda(s)$ is the density at location *s*, *r* is the search radius (bandwidth) of the KDE (only points within *r* are used to estimate $\lambda(s)$, and *k* is the weight of point *i* at a distance of $d_{is}$, which is used to locate *s*.

*iv.* *Buffer:* involves drawing an Euclidean buffer around each facility. Areas within the buffer limits are considered accessible to a facility, while those outside of the limits are considered inaccessible. Areas where overlapping buffers are abundant may represent areas of over service (and vice-versa) [31]. This method can provide a useful access indicator for certain circumstances [32]. For example, this method is useful for assessing the feasible walking range of facilities. However, for many purposes, this method does not sufficiently account for the constraints of availability or transport infrastructure.

*v.* The *Raster calculator:* allows one to create and execute a Map Algebra expression that will output a raster.

# III. RESULTS AND DISCUSSION

## A. Classification and spatial distribution of health care facilities in Yola

Three categories of health care facilities exist in Yola. These categories include *primary, secondary, and tertiary* facilities. Furthermore, these facilities are owned by Federal, State, and Local governments, and by private individuals or organisations.

*i.* *Primary health care:* provision of health care at this level is largely the responsibility of local governments and is supported by the State ministry of health.

*ii.* *Secondary health care:* this level of health care provides specialised services to patients who are referred from the primary health care level. These services are provided through out-patient and in-patient hospital services, which include general medical, surgical, and paediatric cases and community health services. Secondary health care is available at the district, divisional and zone levels of the State. Adequate supportive services, such as laboratory, diagnostics, blood bank, rehabilitation and physiotherapy are also provided.

 ISBN: 978-1-61208-251-6 

*iii.* *Tertiary health care:* consists of highly specialised services, such as orthopaedic, eye, psychiatric, and paediatric cases. These services are provided by education and at specialist hospitals. Appropriate support services are incorporated into the development of these tertiary facilities to provide effective referral services. There are two such facilities in Yola, the Federal Medical Centre (FMC) and the Specialists Hospital.

Generally, maps that depict the locations of health facilities are very useful for health planners. For example, these maps help planners to quickly identify the spatial distribution of health facilities within a community. In addition, these maps allow planners to determine which areas lack such facilities. However, such maps do not exist in Yola. Using *Symbology* from the *Spatial Analyst Tool*, the health care facilities in Yola were mapped based on different classes (i.e., dispensaries, clinics, and general hospitals) and ownership (public and private). Presently, there are 56 health care facilities in Yola, of which 64% are public and 36% are private. Forty of these facilities are located in Jimeta (i.e., the administrative and commercial centre). It can also be observed that the northern and western parts of Jimeta, and southern part of Yola-Town are underserved (Figure 3).

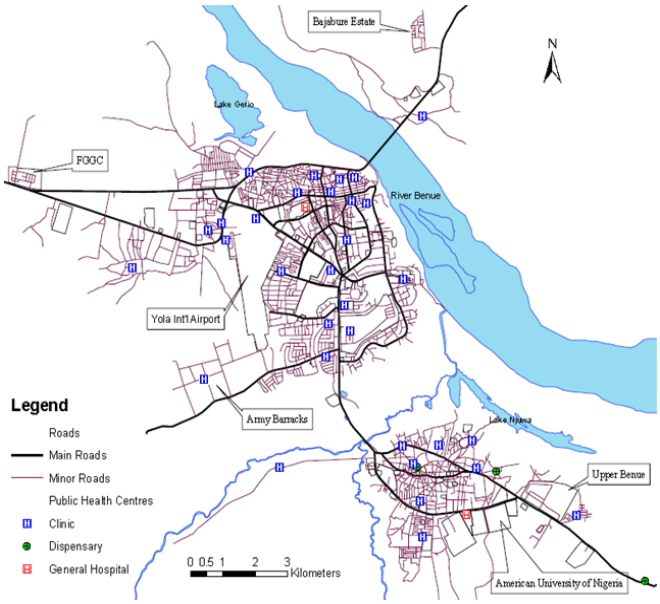



Figure 3. Public health care centres in Yola.

From Figure 4, it can be observed that only three private health care facilities are found in Yola-Town. This means that the residents of Yola-Town have very few options in terms of private health care services or rather they have to travel beyond 10 kilometers to access these facilities in Jimeta. This is quite different for residents of Jimeta that have access within a distance of less than 2,000 m or few kilometers. Though, the majority of the facilities is concentrated in the central part of the city.

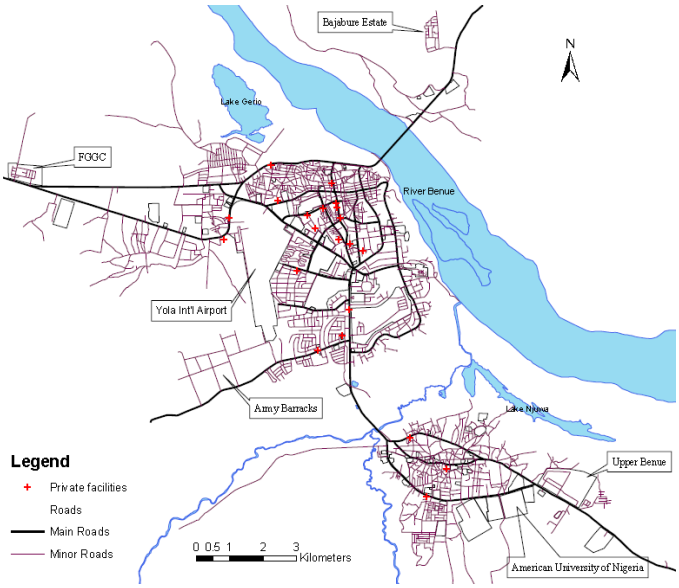



Figure 4. Private health care centres in Yola.

In addition, a KDE was applied by using a search radius of 1,000 m and a cell size output of 100 to determine the density of health care facilities in Yola (Figure 5). The results showed a high concentration of health care facilities in Jimeta, especially in the city centre (i.e., a lopsided distribution pattern which hinders good access to health care service in the city).

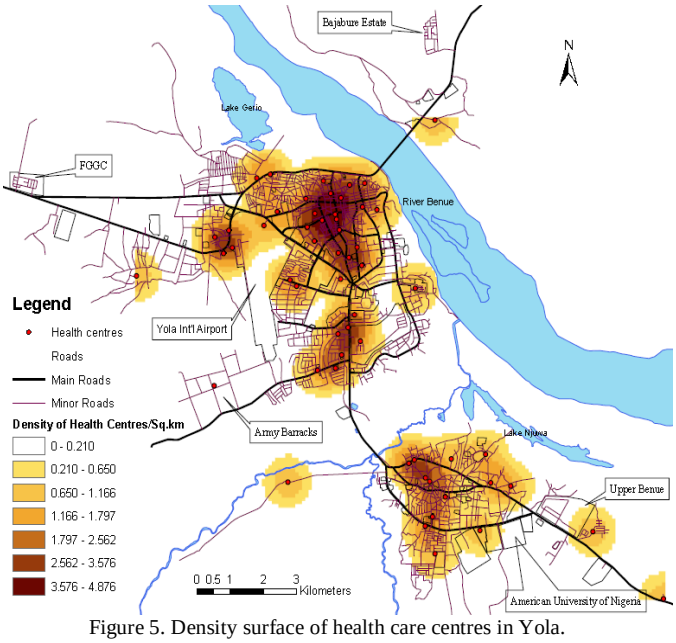



Figure 5. Density surface of health care centres in Yola.

## *B. Population distribution*

Population variables impact the level, nature, type, and variety of health service demands and the resulting service

provisions. Thus, knowledge regarding these demographic variables is essential for the development and delivery of health services because it is used to plan and target service provisions. This observation is particularly true for population based and location based health care planning and community health programs [33, 34].

In an attempt to understand the spatial distribution of the Yola population, the projected 2009 population based on the political ward was used to generate a population density surface by applying the KDE. After several trials, a search radius of 1,500 m with a cell size output of 100 was selected as the most suitable for this application. The result showed that the high density area cover parts of the Rumde, Yelwa, Alkalawa, Luggere, Nassarawo and Doubeli wards in Jimeta. In addition, the Mbamoi, Makama A and B, and Toungo wards were identified in Yola-Town (Figure 6). As expected, these wards are generally high density residential areas, while the low population density areas comprised of medium and low density residential areas. The spatial pattern of the Yola population indicated an increased density in the old city centres and a decreased density in the southern and western areas of Jimeta and in the southern area of Yola-Town. This increasing population density in the old city centres suggests that the demand for health care services in these areas is high relative to other parts of the city.

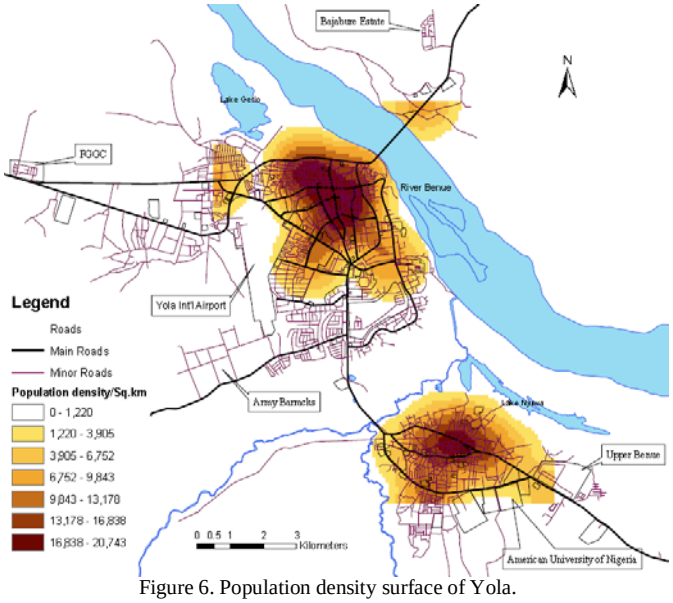



Figure 6. Population density surface of Yola.

## C. *Examining the accessibility of health facilities*

The simplest definition of accessibility for a given location considers how easy it is to get to the location. Defining accessibility to health care centres is an important task for health planners [9]. The concept of health care accessibility is one of the most analysed and debated concepts in public health [35, 36, 37]. The debate on this issue was framed by researchers in terms of economic, structural, infrastructural, and behavioural aspects. This study is concerned with geographic proximity, which is a form of potential spatial accessibility [35, 36].

### *i) Proximity to health facilities*

In this study, a buffer analysis is applied to define the proximity to health care facilities. Buffers were created around all of the facilities in the study area by using the Municipal planning standard with a radius of 1 kilometre in the catchment area of the health care facilities (Figure 7). The results show that some areas in Yola are located outside of the 1 kilometre accessibility zones, especially in the western and northern parts of Jimeta. In addition, it is clear that the existing health care centres must serve a catchment area that is larger than the standard size. Based on this output, different parts of Yola were determined to have low health care accessibility. Health care planners can use this model to make decisions regarding where to build new health centres in Yola. For example, areas beyond the 1,000 m accessibility zones can be used as a reference for determining potential locations for additional health care centres in Yola.

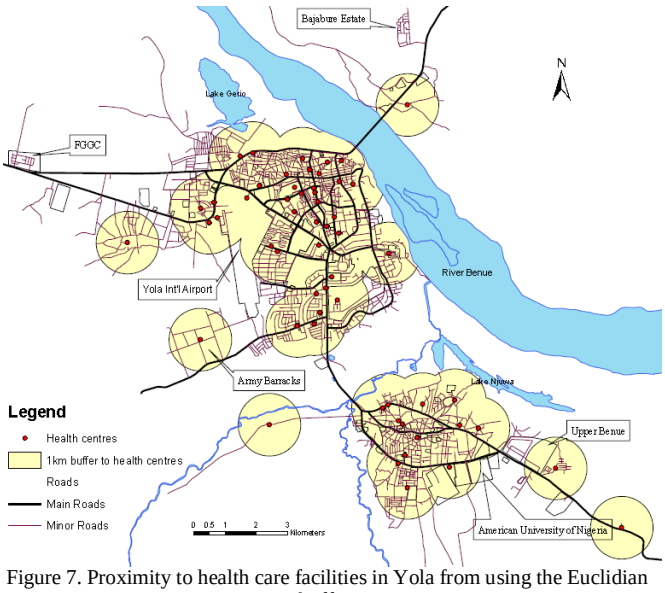



Figure 7. Proximity to health care facilities in Yola from using the Euclidian buffer.

### *ii) Density of physician in Yola*

An adequate supply of well-trained and geographically distributed doctors is critical for providing patients access to high-quality medical care. Moreover, physicians make many of the key decisions regarding the diagnosis and treatment of patients, and they orchestrate the demand for human and other resources in health care [38]. Figure 8 is a KDE map showing the density of physicians based on physician available at each health care centre. Although 71.43% of health care facilities are located in Jimeta, Yola-Town has a higher density of physicians. This distribution pattern is attributed to two factors. First, if a health care centre has a larger number of physicians, then its surrounding areas would have a higher provider density. This scenario was in the case in Yola-Town where the FMC had three times as many available physicians than were available in the Specialist Hospital that is located in Jimeta. Secondly, the

 ISBN: 978-1-61208-251-6 

closer the health centres are to each other, the higher the provider density. This scenario was mainly found in Jimeta, where several health centres are located close to each other that have a comparatively larger supply of physicians than those located in other parts of the city, except for the FMC. In addition, it is important to note that areas with lowest physicians (eastern and western) parts of Jimeta comprised of medium and low density residential areas. On the other hand, it is the high and medium residential densities in the case of Yola-Town.

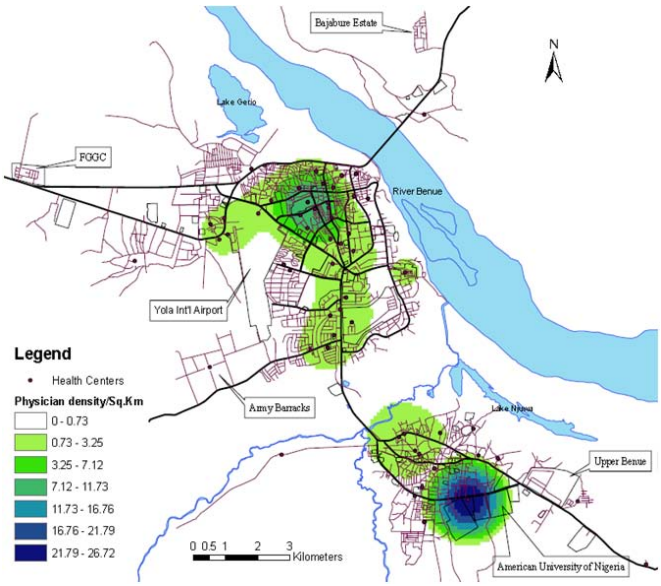



Figure 8. Density of physicians per square kilometre.

## D. *Health centre-to-population ratio*

One way of define the accessibility of health centres is to identify the provider-to-population ratios. This measure is useful for gross supply comparisons between geopolitical units and service areas and is used by policy analysts to set the minimum local supply standards and to identify under-served areas [11]. GIS can be used to compute these ratios by using the KDE, as in [36], or the buffer and overlay functions, as in [8].

The health care centre density and population results (Figures 5 and 6) was analysed by using a raster calculator tool to perform mathematical operations. This type of GIS function was used to assign weights to the raster data for suitability analysis. The population density (in persons per square kilometre) was divided by the density of health care centres per square kilometre. This step was achieved by using the ArcGIS® model builder, which was based on the arithmetic overlay function (Figure 9). The output showed that the majority of in the Yola areas are not well served by health care facilities. This scenario reflects the imbalance of health care facility provisions in the central part of Jimeta and Yola-Town, which have larger population sizes but a limited supply of health care centres. Thus, the existing model(s) that were used by health care planners to providing health care facilities in Yola failed to achieve its ultimate goal. Therefore, several parts of Yola need additional health care facilities. These areas are located in the northern and southern city central districts. This model will be useful if local health authorities plan to increase the supply of health care facilities.

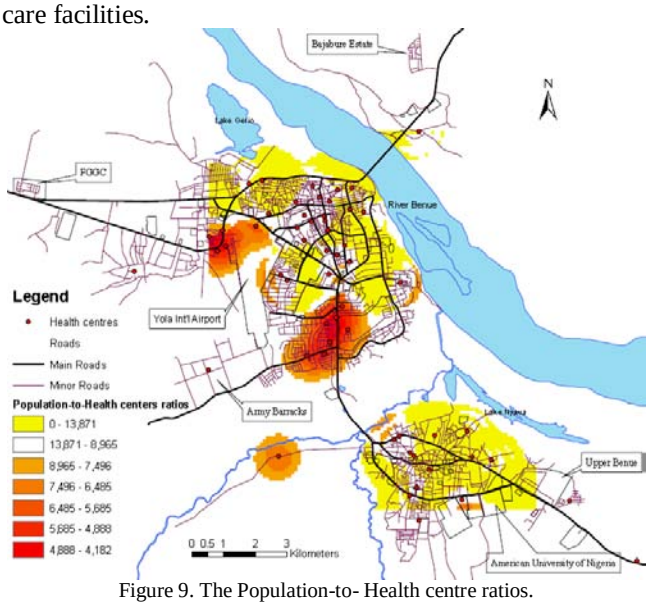



Figure 9. The Population-to- Health centre ratios.

## E. *Physicians-to-population ratio*

For most developing countries, service use and health status data are lacking at an individual level. Thus, the use of the **physician/population ratio** as a proxy to measure the availability of health services and to assess health needs is tested here [38]. Providing the appropriate number of physicians at each location is a prime concern of most health planners, especially when planning the future development of health care systems [39].

Using the population and physician density results (Figures 6 and 8), the same approach that was discussed in section D was applied. The output showed that a gross shortage of physicians occurs in both Jimeta and Yola-Town (Figure 10).

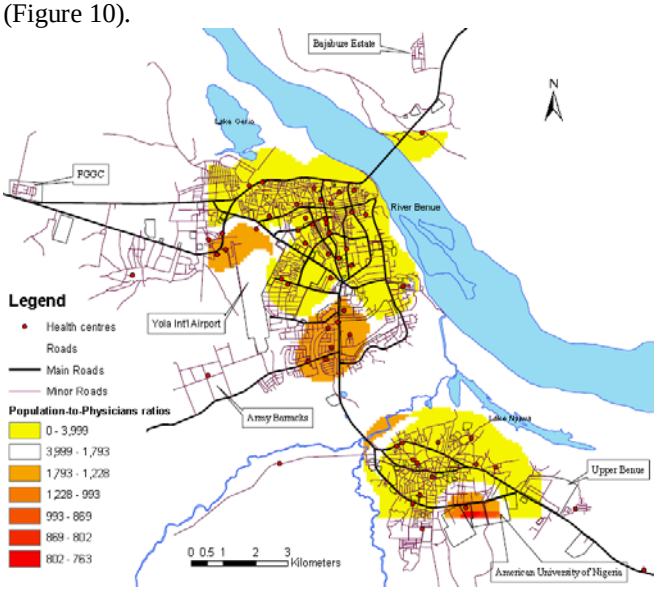



Figure 10. The Population-to-Physicians ratios.

## IV. CONCLUSION AND RECOMMENDATIONS

In this study, we explore the potential use of GIS for modeling the spatial distribution and accessibility of the health care delivery system in Yola. Several digital and non-digital data sets were collected and transformed into GIS data. The GIS tools used in this study included symbols, overlays, KDEs, buffers, and a raster calculator.

Based on the analysis in this study, a gross inadequacy of health care facilities and physicians occurs in Yola. Thus, this study recommends the following: *i)* The urgent improvement of the health care delivery system in Yola for the benefit of the general public should be conducted by providing new health care facilities, upgrading some of the low level health care centres to a higher level, rehabilitating the existing facilities to an international standard, employing more qualified physicians and other supporting personnel, providing incentives for private individuals or organisations to encourage investment in the health sector, and by formulating new policies. *ii)* The adoption of GIS technology by health care planners in Yola.

The study demonstrated that access to spatial GIS information and analysis will help planners make better decisions quicker and at a lower cost than required for the traditional methods. In addition, GIS will enable the establishment of a health care facilities database that can easily be retrieved for analysis at any time.

## ACKNOWLEDGMENTS

A.B. Ismaila wishes to thank The Scientific and Technological Research Council of Turkey (TUBİTAK) for supporting him during his PhD program.

## REFERENCES

[1] United Nations (UN), "Universal Declaration of Human Rights", G.A. res. 217A (III), U.N. Doc A/810, 1948.

[2] O. Nnamuchi, Kleptocracy and its Many Faces: The Challenges of Justifiability of the Right to Health care in Nigeria, *Journal of African Law*, vol. 5, no. 2(1), pp. 1-42, 2008.

[3] K. Prathumchai and L. Samarakoo, 2003, "Applications of GIS in Health care Planning of Lao PDR", http://www.geoinfo.ait.ac.th/research/researchproject.php#2003 (Accessed on 29/09/2012.

[4] A. P. Marks, G. I. Thrall, and M. Arno, "Siting hospitals to provide cost-effective health care", *GeoInfo Systems*,vol. 2, no. 8, pp. 58-66, 1992.

[5] S. J. Walsh, P. H. Page, and W. M. Gesler, "Normative models and health care planning: Network based simulations within a geographic information system environment", *Health Services Research*, vol. 32, no. 2, pp. 243-260, 1997.

[6] M. Zwarenstein, D. Krige and B. Wolff, "The use of a geographic information system for hospital catchment area research in Natal/KwaZulu", *South African Medical Journal*, vol. 80, no. 10, pp. 497-500, 1991.

[7] P. Mitropoulos, and I. Mitropoulos, I. Giannikos, and A. Sissouras, "A biobjective model for the locational planning of hospitals and health centres", *Health care and Management Science*, vol. 9, pp. 171–179, 2006.

[8] W. Luo, "Using a gis-based floating catchment method to assess areas with shortage of physicians", *Health & Place*, vol. 10, pp. 1-11, 2004.

[9] A. A. Murad, A gis application for modeling accessibility to health care centres in Jeddah, Saudi Arabia, In GIS for Health and the Environment Development in the Asia-Pacific Region, edited by P. C. Lai, and S. H. Mak, Springer: Berlin, 2007, pp. 57-70.

[10] F. Wang, and W. Luo, "Assessing spatial and non spatial factors for health care access: towards an integrated approach to defining health professional shortage areas". *Health & Place*, vol. 11, pp. 131-146, 2004.

[11] M. Guagliardo, C. Ronzio, I. Cheung, E. Chacko, and J. Joseph, Physician accessibility: an urban case study of pediatric providers, *Health & Place*, vol. 10, pp. 273-283, 2003.

[12] A. A. Murad, Creating a gis application for health services at Jeddah city, *Computers in Biology and Medicine*, vol. 37, pp. 879-889, 2007.

[13] National Bureau of Statistics, "Annual Abstract of Statistics 2006", Federal Republic of Nigeria, 2006.

[14] A. A. Adebayo and A. L. Tukur, "Adamawa State in Maps", Paraclete Publishers, Yola, Nigeria, 1999.

[15] M. A. Husain, and A. B. Ismaila, "The routines and requirements for the creation of a geographical information system (GIS): The case of GIS database for the Water Board Corporation, Yola", Environ: *Journal of Environmental Studies*, vol. 2, no. 6, pp. 47 -56, 2006.

[16] National Bureau of Statistics, "Directory of health establishments in Nigeria, Federal Republic of Nigeria", 2007.

[17] INEC 2007, Official records, INEC Adamawa State Chapter, Unpublished.

[18] J. K. Wright, 'Crossbreeding' Geographical Quantities, Geographical Review, 45, 52-65, 1955.

[19] ICMOD, http://geoportal.icimod.org/TrainingandEducations/GISforBeginners/p1ch7/ (Accessed on 10/09/2012).

[20] P. A. Burrough, R.A. McDonnell, Principles of geographical information systems: New York, Oxford University Press, 1998.

[21] I. Heywood, S. Cornelius, S. Carver, Introduction to geographical information systems. Pearson Prentice Hall, New York, 2006.

[22] P. A. Longley, M. F. Goodchild, D. J. Maguire, D. W. Rhind, Geographical information systems, West Sussex, England, 2nd edition: John Wiley and Sons, 2005.

[23] Y. Chou, Exploring spatial analysis in geographic information systems. Onward Press, Santa Fe, 1997.

[24] D. O'Sullivan, D. J. Unwin, Geographic information analysis, John Wiley & Sons, Hoboken, New Jersy, USA, 2003.

[25] S. Wise, GIS basics, Boca Raton, FL: CRC Press, 2002.

[26] N. Donthu, R. T. Rust, Estimating geographic customer densities using kernel density estimation, *Markering Science*, vol. 8, no. 2, pp.191-203, 1989.

[27] A. S. Fotheringham, C. Brunsdon, M. E. Charlton, Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons Ltd., Chichester, UK, 2002.

[28] A. Páez, T. Uchida, K. Miyamoto, A general framework for estimation and inference of geographically weighted regression models: 2. spatial association and model specification tests. *Environment and Planning A*, vol. 34, no. 5, pp. 883-904, 2002.

[29] E. D. Yaylaci, A. B. Ismaila, O. Uşkay, Ş. Düzgün, Spatial analyses of electricity supply and consumption in Turkey for effective energy management and policy-making, Energy and

 ISBN: 978-1-61208-251-6 

Natural Resources Policy-Distributional Justace, Springer-Verlag, p.151-166, 2011.

[30] Z. Xie, J. Yan, Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, vol. 32, pp. 396–406, 2008.

[31] J. Spencer, and G. Angeles, "Kernel density estimation as a technique for assessing availability of health services in Nicaragua", *Health Service Outcomes Research Method*, vol. 7, pp. 145–157, 2007.

[32] C. S. Phibbs, and H. S. Luft, "Correlation of travel time on roads versus straight-line distance", *Medical Care Research and Review*, vol. 52, pp. 532-542, 1995.

[33] D. A. Dzewaltowski, P. A. Estabrooks, L. M. Klesges, S. Bull, and R. E. Glasgow, Behavior change intervention research in community settings: how generalizable are the results?, *Health Promotion International*, vol. 19, pp. 235–245, 2004.

[34] D. A. Yeboah, Arab studies quarterly, http://www.britannica.com/bps/additionalcontent/18/25057327/IMPACT-OF-POPULATION-VARIABLES-ON-HEALTH-SERVICES-DEMAND-AND-PROVISION-IN-THE-UNITED-ARAB-EMIRATES (Accessed 05/01/2010), 2007.

[35] A. E. Joseph, D. R. Phillips, "Accessibility and utilization: geographical perspectives on health care delivery", *Harper & Row*, New York, 1984.

[36] M. F. Guagliardo, Spatial accessibility of primary care: concepts, methods and challenges, *International Journal Health Geography*, vol. 3, no. 3, 2004, doi: 10.1186/1476-072X-3-3.

[37] N. Chayovan, A. Hermalin, and J. Knodel, "Measuring accessibility to family planning services in rural Thailand", *Stud. Fam. Planning*, vol. 15, no. 5, pp. 201-211, 1984.

[38] OECD, "Health at a glance 2007 - OECD Indicators, Organisation for Economic Co-operation and Development", 2007.

[39] T. L. Hall, "Why plan human resources for health?", *Human Resources for Health Development Journal*, vol. 2, pp. 77-86, 1998.

 ISBN: 978-1-61208-251-6 

# Object-Based Approach and Tree-Based Ensemble Classifications for Mapping Building Changes

Loubna Elmansouri
Geomatic unity, SURFACES Laboratory
Liege, Belgium
l.elmansouri@student.ulg.ac.be
Hassan II Institute, Department of Topography
Rabat, Morocco
l.elmansouri@iav.ac.ma

***Abstract*-- The aim of this paper is to efficiently detect and identify the building changes from newly registered very high spatial resolution (VHSR) image by comparing with outdated map. The whole process is performed mainly on four steps. First, the image was segmented to generate primitives, which are then represented by a feature vector composed from spectral, geometric, textural and contextual attributes. Thereafter, tree-based ensemble methods (Bagging, Random Forest and Extremely Randomized Trees) are used in a classification step. The final objects' prediction is deducted with respect to the better classifier error rate. Last, a post classification change detection step allows to identify the segments which represent building changes. The data used in this research concerns the city of Rabat (Morocco). A Quickbird image has been used with an old map at the scale of 1:10,000. Regardless of the quality of the detected buildings' shape, the method achieves good rates of completeness and correctness.**

***Keywords; Building changes detection; (VHSR) image; Decision Trees; Random Forest; Extra Trees.***

## I. Introduction

Updated topographic maps are essential in many applications such as land and urban planning. Their update consists on visual interpretation of orthophotos which is time consuming and expert dependent. Recently, some academic studies and cartographic agency experiments present the update process as two automatic steps: (1) change detection and (2) map revision [1,2,3,4]. The first step is the most difficult [4,5]. It can be performed by comparing recent extracted data from a VHSR image with the map to be updated. An object-based approach is more suitable for detecting change. Each extracted polygon is analyzed as a whole allowing more interpretations of change nature. It's almost free data dependent, which ensures its transferability and let it appear as a potential framework for a standard solution of automatic change detection [5]. A typical object-based change detection process relies on two separate stages: (1) persegments-image classification and (2) post classification image-map comparison. First, the image is segmented into groups of spatially connected pixels with respect to some homogeneity criterion [6]. Resulting clumps as the unit of the rest of the process, are more intelligent than individual pixels. They could be described by other features than spectral ones(e.g. geometric and textural) and release better their contextual relationships[1].

Classical classifier cannot handle the high number of features in object-based analysis that's why advanced supervised learning classifiers are becoming more used in VHSR image classification [6]. These include, but are not limited to, decision tree (DT) and tree-based ensemble methods. They are a good data reduction tool and are able to detect multivariate interacting effects between features. They are also non-parametric. No assumption on the distribution of the data is required, it is thus easy to adapt for new datasets. In DT algorithm, a learning set is successively split into binary homogenous subsets based on "if-then" rule tested at each internal node of the tree. The terminal node will be labeled by the majority class. At the end, the tree results in a number of class prediction rules that are used to create a predictive model. Classification accuracy from DT algorithm is often greater compared to using maximum likelihood or linear discriminate function [7]. In change detection context, DT have been used, for example, by Matikainen et al. [8] for mapping urban building changes, by Ruiz et al. [1] to create and to update a high scale database and by Chubey et al. [9] for forest inventories.

The original classification single trees can overfit the training data. It's very much improved by ensemble methods. The tree-based ensemble methods rely on randomization to generate a more stable prediction. These methods are Bagging, Random Forests (RF) and Extra-Trees [10]. Bagging (for "Bootstrap AGGregatING"), builds multiple training subsets by random replacement from the primary learning set. Each subset is inducted as the classical DT algorithm explained above. The Bagging aggregates the predictions of all developed trees by a majority vote to produce the final class. RF combines Bagging with only a random selected subset of k input attributes at each test node. In the Extra-Trees (for "EXTremely RAndomized Trees") method, each tree is build from the complete learning set but at each test node, the best split is determined by randomly selecting one attribute. A complete review can be found in

 ISBN: 978-1-61208-251-6 

[10,11]. Another advantage of these algorithms is their internal faculty to partition learning set into separate sets for training and validation. This allows to automatically determine all their own parameters through cross-validation. Further details about these algorithms will be given thereafter.

Our main objective in this paper is to present a complete automatic procedure for detecting changes. We limit the problem to map building changes which is the most dynamic class in urban and suburban mapping. The proposed methodology is based on a post classification image-map comparison. Tree-based ensemble methods are used in conjunction with an object-based image analysis.

The article is organized as follows: in section 2, study area, data and preprocessing are described. The whole methodology is detailed in section 3. Section 4 provides a discussion of the experimental results that validate the proposed method. Finally, some conclusions are drawn in section 5.

## II. STUDY AREA, DATA AND PREPROCESSING

### A. Study area and Data

The test area is located in Rabat (Morocco) which is a flat relief city. It shows a new arranged residential district which is divided on a regular frame. It's composed of single-family dwellings. Buildings are characterized by different size, shape and construction material.

Our approach was tested on a 1:10.000 scale map and a Quick Bird image. The map has been produced in 1999 from aerial photographs restitution by the National Cartographic Agency of Morocco. The satellite imagery was acquired on August 2004 and has a resolution of 0.61 m in panchromatic band and 2.4 m in the four multispectral ones. Considerable changes have been occurred between the two dates.

### B. Preprocessing

The data pre-processing consists of two major steps: (1) Pan-sharpening multispectral bands by the panchromatic band. The four bands obtained have a higher spatial resolution than the multispectral image and they well preserve the spectral properties from the multispectral image. They will be used in feature extraction stage and (2) as our change detection approach is based on an image–map comparison, precise co-registration between the two data is crucial. Any miss registration can lead to false alarms in detecting change. A polynomial of first order was used; the obtained Root Mean Square Error (RMS) is 0.72m.

## III. METHODOLOGY

The proposed change detection methodology consists in a post classification comparison between a VHSR image and an outdated map. The principal contribution of our methodology is that it integrates advanced image classification within an object-based analysis. Extracted objects are characterized by different attributes (spectral, geometric, textural and contextual). Objects in the map are used to train and to tune parameters in classification learning step. The flowchart of the whole methodology is represented in figure 1. All steps are explained thereafter.

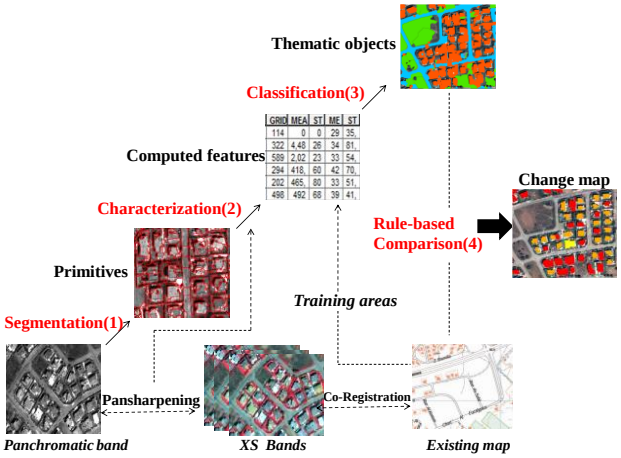



Figure 1. Flowchart of the proposed methodology.

### A. Image segmentation

In this work, a variant of the Watershed Transform (WT) [12] was developed to segment the panchromatic band. The WT holds its foundations from the mathematical morphology. The idea is to consider the image as a topographic surface where the intensity of pixel is considered as a height. Points with divergent flow direction are then seen as crests (objects' contours) which separate catchment basins (homogeneous regions) created from local minima of image (pour points). Usually the WT is calculated from the gradient direction. In this article, the study area concerns the urban context, where the most objects are with almost regular shape, e.g. roads and buildings. That's why a distance map computed from the catchments contours is used to support the WT calculation.

In its original version, the WT produces a profuse number of regions. The small ones come from no-significant local variations of the intensity (false minima). This problem is more marked with the noisy VHSR image. To avoid this over-segmentation, the real minima are to be imposed. All minima in distance map whose depth is less than a fixed threshold are suppressed. We select the threshold automatically according to dynamic range of the distance map.

### B. Features Extraction

Primitives resulting from segmentation were characterized by 4 categories of zonal statistics. Spectral attributes are statistically derived (mean and standard deviation) from 8 spectral and ancillary channels including 4 pansharpened bands and 4 spectral indices (NDVI (Normalized Difference Vegetation Index), SBI (Soil Brightness Index), BAI (Built-Up Areas Index) and SI

 ISBN: 978-1-61208-251-6 

(Shadow Index)) [4]. Geometric features like area, Elongation index and Compactness index were computed. Nine texture features are extracted from the GLCM (Grey Level Co-occurrence Matrix) calculated for all pixels of a primitive, instead of for a regular window size [13]. In our methodology we propose also to take into account the location of the primitive. That's why for each primitive, the all described attributes of its 4 important neighboring primitives were taken as its contextual attributes. In total, 148 features were generated for each primitive.

C. Object-based classification

The classification consists in finding the thematic class of a primitive. After analyzing objects on existing map and available images, we identified 5 classes: buildings, roads, low vegetated areas, urban vegetation and shadow. Objects in the existing map are organized in three themes, buildings, roads and vegetation. The third layer includes urban vegetation (mixels tree-lawn) and low vegetated areas. In our change detection purpose we look for the large low vegetated areas that could have changed to build up areas. However, no information concerning vegetation type is associated to exiting vegetation objects. Whereas such knowledge is necessary to use map objects as training data. Nevertheless this information can be obtained from image data. On the other hand, shadow segments could not be trained from existing map. In order to cope with this, the classification is performed in a two pass scheme: (1) urban vegetation and shadow discrimination and (2) Tree-based ensemble classifications.

First the mean value of the NDVI layer was used to create a binary mask. Two classes were defined, urban vegetation and non-urban vegetation. The latter class was further subdivided into shadow and non-shadow classes. This separation is done by thresholding the SI layer. Shadow segments are the darker in the image, so it's assumed that they have the lower SI value. The non-shadowed primitives are then classified in 3 themes: buildings, roads and low vegetated areas thanks to tree-based ensembles methods. The training samples are automatically selected from the old map. For every object of each layer in the old map, the segment that corresponds spatially to it is identified on the segmented image. The primitive is a priori retained for training if its geometric properties (shape and area) are close from those of the map object. Specific class rules could be added at selection stage. For example, having a large area is an additional criterion for selecting segments to train low vegetated class. Resulting segments attributes data were used as a learning set to generate tree-based ensemble models.

We have decided to employ tree-based ensemble methods for classification for many reasons.

(1) They produce multiple models which will feign the variability of real data.

(2) They have been successfully applied in a variety of pattern recognition and they are often unexcelled in accuracy among current supervised learning algorithm (e.g. Support Vector Machine [7]).

(3) They have been used increasingly in recent years in VHSR image classification and with better results than other classification approaches [14]. For example, for determining agriculture management practices [14], as well as for land use/land cover mapping change projects [15].

Many statistical packages Pepito [16], R, S-Plus [14], CART [6] and Matlab [17] propose a complete toolbox for implementing Bagging and RF. The developed codes in [11] include also the Extra trees algorithm; they were used to implement our classifications. Two parameters are to fix, splitting criterion and stop splitting criterion. To split a node, a score measure is defined by minimizing the impurity of the output variable in the local subset (which also could be expressed by maximizing the class separation at each node). The GINI's index [11] is used. For the second criterion, It's commonly used that de development of a branch stops when the number of instance is lower than a given threshold or when impurity cannot be further reduced (when the output variable is constant). With the exception of k the number of a randomly selected attributes in RF which is fixed at the root square of the total number of attributes, all other parameters (number of developed trees(T) and stop splitting criterion (Nmin)) were adjusting by cross validation from learning set according to the best error rate.

D. Rules-based change detection

Changes detection are looked for by a post-classification comparison of the existing building layer with the building classification results, which is converted without generalization to a vector format. For comparison, each polygon must be matched with its counterpart in the other dataset. The obtained positional accuracy of the map and image is so good that matching of polygons can be based on their overlap. However, when defining rules for detection change, thresholds must be large enough to tolerate small errors in location and shape of buildings. These errors result mainly from (1) attributes errors propagation and (2) intrinsic difference in the nature of objects in both datasets (map and image). According to the ratio of overlap area (A_overlap) between a pair of buildings on the map(A_map) and in the building detection ( A_detection) result, the correspondence between polygons is performed. We have defined 5 classes for buildings change:

**N**ew building: no building on the map corresponds to the detected building with:

$$A\_overlap / A\_detection < Thr1 \quad (1)$$

**D**emolished building: no building on the building detection corresponds to the building in the map with:

$$A\_overlap / A\_map < Thr1 \quad (2)$$

**C**onfirmed existing building: building on the map corresponds to one in the building detection with:

$$A\_overlap / A\_map > Thr2 \quad (3)$$
and
$$A\_overlap / A\_detection > Thr2 \quad (4)$$

**E**nlarged building: building on the map corresponds to one in the building detection with:

$$A\_overlap / A\_map > Thr2 \quad (5)$$
and
$$A\_overlap / A\_detection < Thr2 \quad (6)$$

**R**e-examination needed: includes all other cases where none of the aforesaid conditions is verified.

In our case, Thr1 is fixed to 10% in order to take into account the accumulated errors throughout the change detection process. For the second threshold Thr2, Rutzinger et al. [18] suggested that a threshold value between 50% and 70% should be selected for this type of comparison. In [19], the authors propose to select the threshold of 80% for industrial areas, 70% for the apartment house areas and 60% for areas with smaller detached houses. We select Thr2= 60% because our first purpose is to detect all unregistered buildings without caring too much about the quality of their shape.

## IV. EXPERIMENT RESULTS, EVALUATION AND DISCUSSIONS

### A. Classification

First of all, aforesaid classification algorithms were evaluated. In an object-based classification method, the unit for classification is the object. Consequently, the sampling unit for validation also has to be an object. 99 segments selected by a random stratified sampling from the no learning data and annotated visually from the image by an independent interpreter were used as validation data. The overall accuracy, the coefficient kappa and its standard deviation were calculated [20]. Table I presents the best result obtained for each classifier.

TABLE I. CLASSIFICATION ACCURACY FOR ALL EXPERIMENTED CLASSIFICATION ALGORITHMS.

| Algorithm | Overall accuracy % | Kappa % | $\sigma_{Kappa}$ |
|---|---|---|---|
| **Classical DT** | 59.18 | 0.50 | 1.1 |
| **Bagging (T=50)** | 73.56 | 0.64 | 0.11 |
| **RF (T=200)** | 82.05 | 0.72 | 0.04 |
| **Extra Trees((T=500)** | 82.8 | 0.73 | 0.02 |

As expected, the results from classical DT are not good what is due to the high variance of this method, moreover that no pruning technique was used. Note that the three scores of tree-based ensembles methods are better than the classical DT (14% at list). RF and Extra trees present better scores than Bagging with a light superiority of the Extra trees. The RF incorporates a features selection process that's why it maintains an increase even if with a high number of attributes. Extra trees builds completely random independent trees, the constructed model lacks sensitivity to noise and is not subject to overfitting, which explain the stability of its results ($\sigma_{Kappa=}0.02$).

For the rest of the process, the predictions from Extra trees will be adopted. Table II shows the error matrix. A good separation between "roads" and "buildings" is achieved. The major problem in the classification result is the asymmetric confusion between "buildings" and "low vegetated" segments, which is the origin of the small kappa's value. Adjacent segments to the buildings are mixture of grass, trees and asphalted surfaces; they are either court yards or terraces. The classification algorithm labels them as buildings because of their close spectral response to building. Moreover on the one hand, their area is small and on the other hand, the image acquisition date (August) is not suitable for separating low vegetated and build-up areas in Morocco. The classification result is shown in figure 2.

TABLE II.EXTRA TREES CLASSIFICATION CONFUSION MATRIX (B: BUILDING, R: ROAD, LV: LOW VEGETATED).

| | | Reference | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **B** | **R** | **LV** | **Total line** | **Prod-ucer's%** | **Omiss-ion %** | **Over-all %** | **Error rate %** | **Kappa** |
| **Classification** | **B** | 44 | 1 | 0 | 45 | 97,8 | 2,2 | 82,8 | 17,2 | 0,73 |
| | **R** | 3 | 21 | 0 | 24 | 87,5 | 12,5 | | | |
| | **LV** | 12 | 1 | 17 | 30 | 56,7 | 43,3 | | | |
| | **Total Column** | 59 | 23 | 17 | 99 | 80,6 | | | | |
| | **User's %** | 74,6 | 91,3 | 100 | 88,6 | | | | | |
| | **Commission %** | 25,4 | 8,7 | 0 | | | | | | |

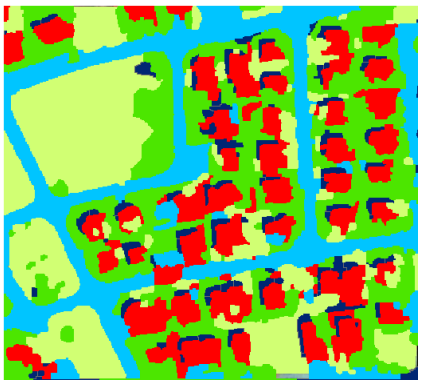


Figure 2. Classification result ( building road low vegetated shadow urban vegetation).

B. Change detection

Applying change detection rules directly to the buildings class leads to an overestimation in changes. A series of contextual rules were then defined so as to optimize the building classification result. They are based on the fact that the building cast shadow in the opposite direction of the sun. First, for each shadow segment, its neighboring segments in building class with the closest azimuth to the sun's were selected. Once confirmed as a construction segments, an homogeneity analysis is then performed to detect the construction segments from their neighboring segments among the remaining ones in the building class. We run this last step iteratively until no neighboring is detected yet. Segments which are not selected at any iteration were excluded from the further process.

To evaluate change detection results, a reference change map was performed by an exhaustive computer assisted image interpretation using the image and the old map as backdrops. Buildings in the same class in reference change map that touch each other in space were considered as one building. The concept and parameters of the change detection validation scheme were similar to the ones defined for the change detection stage. If there's any overlap higher than 60% between buildings of a pair on our change detection result and in the reference change map and which are labeled by the same class in both datasets, this is considered as a good detection. However, when the two counterpart objects are labeled separately, the detection is correct but with a wrong identification of change nature. Both automatic change detection result and reference change map are presented in figure 3, the confusion matrix of the change detection validation is given in table III.

(a)

(b)

(c)

Figure 3. (a) Building layer in 1999,(b) Our change detection map, (c) Reference change map ( New B. Confirmed Existing B. Enlarged B. Re-examination needed).

TABLE III. CHANGE DETECTION CONFUSION MATRIX

| | | Reference | | | | |
|---|---|---|---|---|---|---|
| | | N B. | C B. | E B. | D B. | Total |
| Change detection | New B. | 23 | 0 | 0 | 0 | 23 |
| | Confirmed Existing B. | 0 | 22 | 0 | 0 | 22 |
| | Enlarged B. | 0 | 01 | 0 | 0 | 01 |
| | Demolished B. | 0 | 01 | 0 | 0 | 01 |
| | Re-examination needed | 0 | 02 | 0 | 0 | 02 |
| | Total | 23 | 26 | 0 | 0 | 49 |

The comparison of the change detection result of our automatic procedure with the reference change map shows that 46 buildings from the 49 ones present on the area are detected so as 94% for completeness rate. The change nature for 45 buildings of the detected ones is correctly identified so as 92% of correctness rate. Only one detected building has a wrong class of change. It concerns an unchanged existing building which is labeled as enlarged one by our approach. This is due to the segmentation step, as we merge small segments with their adjacent ones to overcome the over-segmentation problem, some segments become larger. Omission error is 2%, only one building is completely not detected (demolished). It was classified as a road segment. This miss-classification is explained by the selected stopping criterion (Nmin=1) which forced the branch to continue split until only one segment per class. Two existing buildings were detected with smaller surface area than in nature, which doesn't allow verifying the conditions to be confirmed as existing buildings. Indeed, the missing parts were first classified as buildings, but as they are composed of mixed pixels (build-up areas and vegetation) they present a spectral heterogeneity with their neighboring which leads to their elimination by homogeneity post classification rules. Further attention is needed to this change detection class. Note that no false alarm was indicated.

To have geometric information about detected changes, an area ratio was calculated for each

 ISBN: 978-1-61208-251-6 

detected building. It presents the object surface area in reference map which is covered by the object at the change detection result of our automatic approach. The mean ratio of detected area per building is 85%. More geometric evaluation is planned in the further developments.

## V. CONCLUSION AND FURTHER DEVELOPMENT

An automatic change detection process was proposed to map building changes. The first obtained results (94% for completeness and 92% for correctness) are promising and are comparable to the best results for similar researches found in literature (98% for completeness and 96% for correctness) [5]. The whole process is conceived in four separate stages. The basic concepts for each stage are chosen to be opened and generic. So that the minimum of adaptations will be required when the process is applied in a different context.

Completeness is a decisive indicator since it is important to be exhaustive, to be sure not to miss any changes. Searching for undetected changes is more difficult and contrary to the basic goal of using automated change detection. For new buildings class, it's completely correctly detected which might lead to an operational use of our procedure. The result of this class could be used to trigger and to guide on-site inspections by the control authority. Unchanged existing buildings class represents 53% of the total number of buildings in the test area. This class is correctly detected with 85% rate. In theory, when updating the map, operator could bypass this class and concentrate on checking the changed ones. Much time and costs might be saved.

Up to now, our main goal was to develop a complete automatic procedure for mapping buildings changes. Once ready, additional studies are underway to show if the procedure is transferable for other contexts. Different datasets with heterogeneous areas are planned to be tested. Practical tests are also needed to show if the results are geometrically accurate enough for an operational use.

To update the old map, the changed buildings must integrate it. A shape adjustment step appears necessary to correct their contours before. Developing snakes from the detected building contours to create corrected ones seems to be attractive.

## REFERENCES


[1] L.A. Ruiz and T. Hermozilla,"A multi approach and object-based stratey for updating lu/lc geo-databases based on Worldview-2 imagery", 2011.

[2] V. Walter, "Object-based classification of remote sensing data for change detection". ISPRS Journal of Photogrammetry & Remote Sens., 2004, pp. 225-238.

[3] D.A. Holland, D.S. Boyd and P. Marshall," Updating topographic mapping in Great Britain using imagery from high-resolution satellite sensors". ISPRS Journal of Photogrammetry and Remote Sensing 60 (3), 2006, pp. 212-223.

[4] M. Bouziani, K. Goïta and D.C. He, "Change detection of buildings in urban environment from high spatial resolution satellite images using existing cartographic data and prior knowledge". In: Proc. IEEE International Geoscience and Remote Sensing Symposium, Barcelona, July 2010, pp. 2581-2584.

[5] N. Champion, "Detection of Unregistered 2D DataBases". European Spatial Data Research report, 2010.

[6] R.L. Lawrence and A. Wright, "Rule-based classification systems using classification and regression trees (CART) analysis", Photogrammetric Engineering & Remote Sen., 2001, pp. 1137-1142.

[7] M.A. Friedl and C.E. Brodley, "Decision tree classification of land cover from remotely sensed data", Remote Sens. Environ, 61,1997, pp. 399-409.

[8] L. Matikainen, J. Hyyppä, E. Ahokas, L. Markelin and H. Kaartinen, "Automatic Detection of Buildings and Changes in Buildings for Updating of Maps", Remote Sens. 2010, pp. 1217-1248.

[9] M.S. Chubey, s.e. Franklin and M.A Wulder, "Object-based analysis of IKONOS-2 imageryfor extraction of forest inventory parameters". Photogrammetric Engineering and Remote Sensing, 2006,72, pp. 383-394.

[10] R. Maree, "Decision trees for automatic images classification". Phd Thesese. Ulg university, 2005.

[11] P. Geurts, A. Irrthum and L. Wehenkel, "Supervised learning with decision tree-based methods in computational and systems biology". Mol Biosyst 5, 2009, pp. 1593–605.

[12] B.T.M. Jos and A. Meijster," The Watershed Transform: Definitions, Algorithms and Parallelization Strategies". Fundamenta Informaticae. 2001.

[13] R. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification". IEEE Transactions on Systems, Man and Cybernetics, 3(1),1973, pp.610-621.

[14] J. D. Watts and R. L. Lawrence, "Merging Random Forest classification with an object-oriented approach for analysis of Agricultural Aands". ISPRS Journal of Photogrammetry and Remote Sensing, Working Group VII/4.2010.

[15] C. Leignel, O. Caelen and O. Debeir, "Detecting man-made structure changes to assist geographic data producers in planning their update strategy". ISPRS Journal of Photogrammetry and Remote Sensing, 2010, part,4-8-2-w9.

[16] www.pepite.be

[17] Matlab (2010). Matlab documentation: Image processingtoolbox. http://www.mathworks.com/help/toolbox/images)

[18] M. Rutzinger, F. Rottensteiner, N. Pfeifer, "A comparison of evaluation techniques for building extraction from airborne laser scanning". IEEE J. Selected Topics Appl. Earth Obs. Remote Sens. 2009, 2, pp,11-20.

[19] L. Matikainen, L. Hyyppä and H. Kaartinen, "Automatic detection of changes from laser scanner and aerial image data for updating building maps. Remote Sensing". The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2004,XXXV/B2, pp.434-439.

[20] R.G. Congalton, "Accuracy assessment of remotely sensed data: future needs and directions". In: Proceedings of Pecora 12 land information from space-based systems. 1994, pp. 383- 388.


 ISBN: 978-1-61208-251-6 

# XGIS Flex: A Free and Open Source Framework for the Development of the Geographic Information Systems on the Web

Mariano F. Pascaul, George Sand L. Araújo
Geosciences Institute
University of Brasilia
Camp. Darcy Ribeiro –DF– Brazil
{mariano, georgesand}@unb.br

Maristela Holanda
Depart. of Computer Science
University of Brasília
Camp. Darcy Ribeiro –DF– Brazil
mholanda@unb.br

Edson E. Sano
Inst. of the Environment and
Renewable Natural Resources
SCEN trecho 2 Ed.Sede – DF – Brazil
sano@ibama.gov

***Abstract*—Geographic Information Systems on the web (web GIS) are currently implemented for different purposes. This paper presents a framework to develop web GIS with free and open source software, for interoperability between different sources of heterogeneous data, as well as the visualization of maps in different formats with components implemented with a Web 2.0 concept. We present two case studies that implemented a web GIS to Japan International Cooperation Agency – JICA together with the Secretary of Planning and Modernization of Public Management by the Government of the State of Tocantins – Seplan-TO and a web GIS for the Brazilian Institute of the Environment and Renewable Natural Resources – IBAMA.**

***Keywords-GIS; web; Flex; open source***

## I. INTRODUCTION

A Geographic Information System (GIS) is frequently defined as the combination of a database management system, a set of operations for exploring data, and a graphic display system that is used for geospatial analysis. These GIS analyses have the main purpose of supporting decision making and modeling some of the possible consequences of those decisions [2][4][7][27][30]. GIS environments are also cartographic tools that facilitate creating maps and examining the impact of changes to the maps interactively [14] [17][21][26][27].

GIS on the web (web GIS) is currently being developed, and one challenge in that environment is interoperability among heterogeneous databases. For interoperability of the data, web services technology is being used [10]. The standard set by the Open GIS Consortium (OGC) proposes open service architecture of web GIS to support data-interoperability. And, it suggests the use of Geographic Markup Language (GML) based on XML to exchange the data between the web client and the web GIS [13][16]. The REST [22] technology is also used to support interoperability with geographical databases.

For the visualization of maps in an interactive way, Web 2.0 technology is being applied through different components of Rich Internet Application (RIA) [35]. As is observed in [15], this technology is being applied in the development of web GIS. Web mapping applications such as Google Maps, Google Earth, Microsoft Bing Maps and Yahoo Maps are usually considered good examples of Web 2.0 [10].

In this environment, the free and open source software FOSS community has developed solutions for GIS implementation, database management systems with free spatial extension and high performing map servers. One of the elements that facilitated the distribution and use of these free tools was the standard defined for interoperability between data and the applications used for the development of a GIS [6] [29].

This paper presents XGIS Flex — a framework to develop web GIS with free and open source software, with interoperability between different sources of heterogeneous data, as well as the visualization of maps in different formats whose components are implemented with a Web 2.0 concept. This framework was used to implement two case studies presented in this paper.

The content of this paper is divided into the following sections: 2 – basic concepts about web GIS are presented; 3 – the XGIS Flex framework is defined; 4 – Related Works, analyses; 5 – Case Study, where the architecture was used to develop the GIS for the Japan International Cooperation Agency – JICA in partnership with the Secretary of Planning and Modernization of Public Management by the Government of the State of Tocantins – Seplan-TO and of the Brazilian Institute for the Environment and Renewable Resources – Ibama; and finally, 6 – Conclusions.

## II. GEOGRAPHIC INFORMATION SYSTEM ON THE WEB

Geographic Information System on the Web is any GIS that uses web technologies. The simplest form of web GIS should have at least a server and a client, where the server is a Web application server, and the client is a web browser, a desktop application, or a mobile application [5][10].

The architecture of a web GIS is most commonly based on the use of three layers: User Interface Layer, Application Server Layer and Database Layer [3][9][28]. Some authors considered four layers, where the integration layer is added on the web GIS architecture, which is based on web services [32].

 ISBN: 978-1-61208-251-6

The User Interface layer serves as a graphic user interface (GUI) to present the result of spatial data, allowing the end users to interact with the backend services. The Application Server layer communicates with multiple data sources via the data integration layer, and interacts with end users to analyze and manipulate data coming from data provider services. The Database layer of data provider services is a set of remote data provider services for data sharing. Each data provider service offers a set of interfaces through which client applications can pull remote data in and manipulate the data.

## III. XGIS FLEX FRAMEWORK

The XGIS framework was developed to be a free and open source option for the platforms developed in Flex. The XGIS enables the navigator for use in RIA and high performance Web 2.0 concept environments. There is interaction with OGC standard web services, such as WMS, WFS, WFS-T e WCS, REST, as well as files such as GPX, KML, SHP and other formats including JSON and GeoRSS. With XGIS, it is possible to visualize data from different web services. It is not aim of the XGIS to offer web services. Besides visualization, the XGIS has some functionality, such as, extract data from many formats, measure distances, area and several graphics. In the next section, each one of these functions is shown through the components.

The XGIS is strongly linked and integrated to the Map Server Geoserver (Java), which is a high performance interactive map server as presented in [25], having interoperability of the open source with multiple connectors for the sources of data such as PostGIS, Oracle, SQL Server, DB2, SDE, shapefiles and others. With this framework it is also possible to visualize data coming from other map servers, such as ArcGIS Server or MapServer and fuse services, such as, Microsoft Bing, OpenStreet Maps (OSM), Yahoo Maps and others.

### *A. XGIS Flex Framework Components*

The XGIS Framework has various components. These components were developed based on the architecture presented in [23]. Figure 1 presents the components, where the XGIS framework components are represented by strong lines, and dashed lines are presented for the software used in the web GIS environment. Each one of the components of the framework is described below.

The GPX, KML and Shapefile Overlay Component enables the opening of shapefile files and their attributes, providing the overlay of geometries available in .shp files in the web Flex environment. The constant attributes in .dbf files are presented in a tabular format and related to the correspondent geometry. The features that can be viewed are: line, point, or polygon. If the set of files is projected in a system distinct to the map base, the inclusion of the compacted file .prj is necessary for the system to complete re-projection. This component reaches a large quantity of Garmim brand GPS field equipment users and, therefore, this was strongly recommended in the development of this component, which makes the overlay of files downloaded directly from equipment of this type possible. All of the waypoints available to track overlays of a line are presented, in which the initial point is presented in green, the path in yellow and the end in red. The attributes given to the paths are shown in a floating table. The KML Overlay is a component that provides the opening and comparison of the routes generated in Google Map platforms, and also in simple files in KML format; the available attributes in the file referenced are also presented in splints through floating panels.

The Vectorization component class serves the vectorization and release in distinct database formats, such as lines, points and polygons. Algorithms for the calculation of the area and perimeter are also used in the class; the user can opt to save the designed features on a local disc in JSON format or directly in the database. When drawing, it is possible to add or edit new vertexes and exclude former ones or all polygons, lines or points drawn.

The Export Data Component is totally integrated with Geoserver Java map server. Through this component, it is also possible to make the selection and later export and download layers or a part of them in diverse formats such as shapefile, csv, pdf, xls, kml, kmz, jpg, png.

The Navigation and Data Tabs Component is a presentation component, in which formatting grids were built with the aim of improving the visual aspect of register lists extracted from the database. There is frequent interaction between the navigation and consultation of attributes related to the geometry visualized.

The Treatment of Geographic Database Component aims at adopting the tool of a set of classes capable of treating incoming information from the interface, which will be sent to the database. With this tool one can select which type of database will be used and the classes will be the interactions necessary for a correct treatment of distinct types between the manufacturers of DBMSs. Treatment of geographic features is a set with the minimum rules necessary to avoid classical errors at the moment of vectorization, such as the creation of polygonal ties.

The Charts are tools for which geographic database tables, or simply web services about them, can be chosen by way of selection boxes, from which one can select numerical attributes and vectorize an area on the screen resulting in charts. These are interesting analytical tools for statistical data, however with geographic presentation and selection. Taking advantage of the available functionality in the SDK Flex, the following types of graphics were used: lines, columns, bars, pizza, candlestick and areas.

Media Overlay is the component responsible for the overlaying of videos, photography and documents over maps, for the most part, representing a strong Web 2.0 concept. For the development of this component, it was necessary to map and treat the types of media data available in the main DBMSs on the market so that the consultation and overlaying of these on the maps become possible.

### *B. XGIS to Develop web GIS*

The XGIS framework was designed to be user friendly. If the user of the XGIS is not a software developer, he or she

 ISBN: 978-1-61208-251-6 

will still be able to customize an application easily and quickly, needing only to edit XML extension files. With the configuration files in XML one can define which map services can be used, in which tables it will be possible to edit data, titles and subtitles of the system, logotypes or brands, data subtitle, symbols, and the individual configuration of each component developed.

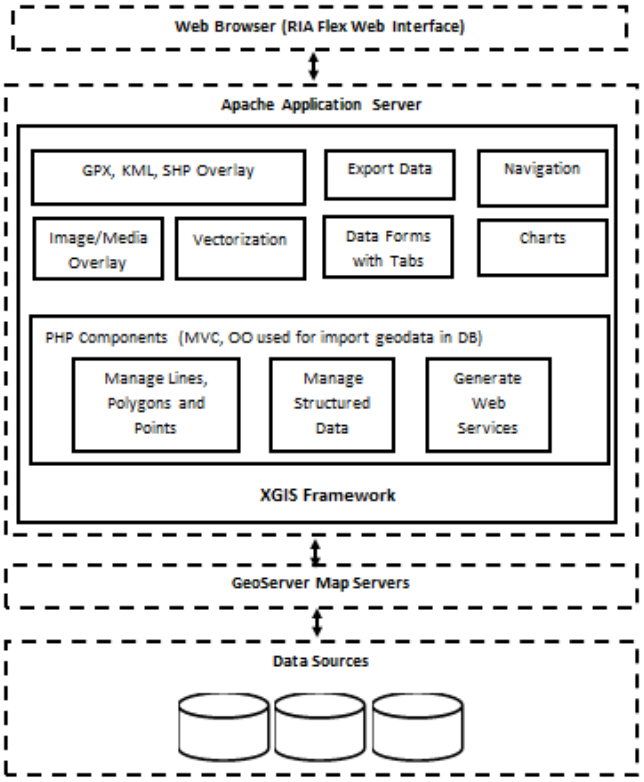



Figure 1 – Components of the XGIS.

Figure 2 lists a piece of the code of the XML file of the configuration of layers to be shown on the spatial interface with tags already configured.

```
<?xml version="1.0" encoding="UTF-8"?>
<camadas>
  <camada>
        <name>Estados</name>
        <url>http://corredorjalapao.seplan.to.gov.br:8080/geoserver/jalapao/wms</url>
        <layers>jalapao:estados</layers>
        <projection>EPSG:900913</projection>
        <format>image/png</format>
        <maxExtent>-180,-90,180,90</maxExtent>
        <version>1.1.0</version>
        <visible>true</visible>
        <transparent>true</transparent>
        <tiled>true</tiled>
  </camada>
  <camada>
        ...
  </camada>
        ...
</camadas>
```

Figure 2 – XGIS Configuration File.

Besides being able to quickly implement an interface of interactive maps, another feature of the framework is in the engine of web service for the interaction between the web forms projected in the Flex environment and different types of databases. The engine was built in PHP language and treats connections with the SGBDs MySQL, PostgreSQL, SQLite, Oracle and SQL Server and its spatial extensions, thus, it is possible to have a programming language for developing new components that interact with textual and geographic data. In case the user of the Framework does not want to use the PHP motor incorporated, he can use whichever programming language he wants to Interact with the maps: C#, Java, Phyton, among others. The tools just need access to the data via web-services that can be easily available in other languages.

The layer of data presentation always exhibits a standardized layout, as shown in Figure 3, making the use of the application easier. This characteristic is possible because the components that were developed in Flex using ActionScript follow the MVC—Model, Vision and Control model [36]. The developer has only to design the form using a graphic environment and make it available to the web service corresponding to the database tables he or she wishes to interact with using and extending the set of codes developed. It does not matter which DBMS the user uses, it's merely necessary to alter the data that pertains to the connection, such as user, password, IP address and type of DBMS to be connected.

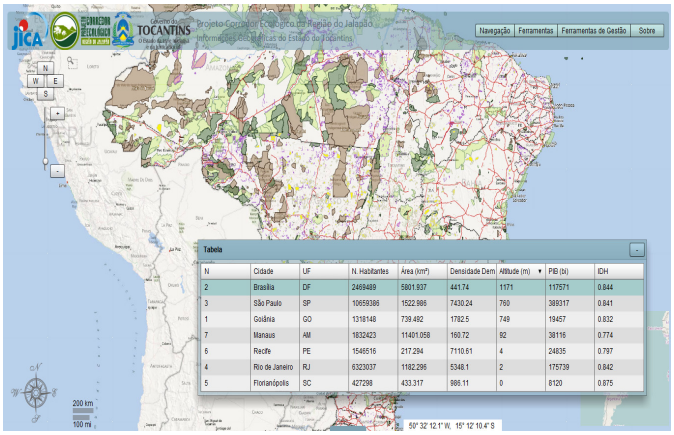



Figure 3 – XGIS Standard Layout.

## IMPLEMENTATION OF THE XGIS FRAMEWORK

Some free and open source libraries were used for the development of the XGIS components. The highlighted libraries described below are among the libraries used.

The SDK Flex [1] has a set of components available, including: buttons, list boxes, tree-views, data grids, text control objects, layout containers, graphic components, animations, status applications, modal dialogue boxes, among others. The Openscales libraries [18] were also widely used and an extension of the majority of the classes of the library, as well as the creation of new components and classes.

Other sources and free components written in ActionScript were used and organized within the XGIS

 ISBN: 978-1-61208-251-6 

framework, such as, file reading components of extension shp, shx and dbfs and, of the watermark insertion in the application [31].

The PHP components manage the entrance of vectoral data that are sent by web interface such as lines, point, polygons and projection systems. Textual data are also controlled, making not only the presentation of information over maps easier, but also the insertion, alteration, exclusion and consultation of the same directly from the data base.

## IV. Related Works

Different tools for aiding in the development of web GIS with free and open source software are found in the literature and there are proprietary tools with closed sources. Some of these tools are presented as follows.

The ESRI makes an API for Flex [8] available for developing web GIS with web 2.0 characteristics, however, it is a proprietary closed source framework. A negative point is the strong linkage existing between the ESRI framework and map services originating from ArcGIS Server servers, also proprietary and costly to implement.

Google has an environment available for development in Flex, Google Maps API for Flex [12], however, it is also proprietary. The great advantage of this environment is the popularity of its base maps, which have excellent quality since Google invests heavily in the acquisition of cartographic bases, and the service provides very high performance as well. The negative points are the low quantity of available functions, presenting itself more as a compiled library to be incorporated to another Flex application and its proprietary source is unalterable.

The i3Geo [24] is a web software based on a set of other free software, mainly Mapserver. The main focus is the availability of geographic data and a set of tools for navigation, analysis management, on demand map sharing, and management. It is strongly linked to MapServer, free standard web services management software OGC, but from the first generation not high performing nor created for editing geographic data via web, as is Geoserver.

Four libraries are considered free frameworks for web GIS development written in JavaScript: OpenLayers [19], GeoExt [11], OpenScale [18] and MapFish [20]. Each one of these is described as follows.

OpenLayers is a JavaScript library accessed as API, however, with an open source and one that currently presents as a library used mainly for the support and inclusion of maps on web pages, this is because of its versatility, able to adapt to diverse languages and modern web browsers. Another advantage is in the capacity of mash-ups between sources such as Google Maps, MsBingMaps, Yahoo Maps, OSM and others.

GeoExt presents itself as a JavaScritp Toolkit for Rich Web Mapping Applications that is basically an OpenLayers library, but with resources that raise the level of the applications developed, making them closer to a desktop application, while still in a web environment. It is a recent project, still in its first versions.

Openscales is a framework written in ActionScript 3 with an LGPL license type. Its libraries are widely used in our solution, but are limited to support the navigation maps. Classes used in the framework presented in this paper have been extended, the engine optimized and many additional components were created.

MapFish is a framework to be developed from the web applications written in Phyton. These new classes present functionality for the construction of applications of the web mapping type. MapFish makes a Java Script toolbook available composed of Openlayers, GeoExt and ExtIS.

Unlike the libraries listed above, the XGIS is a full framework for developing web GIS since it interacts manipulating textual and spatial data from different DBMSs, besides dealing with the rendering of maps and pictures, and interacts as a tool Web 2.0 RIA with photos, georeferenced video and documents. The main idea of XGIS is the possibility of accelerating the development of GIS applications for web environments that are easily integrated and do not require programming knowledge to implement, and always maintain a good visual presentation.

## V. Case Study

The XGIS framework was used in different GIS, as we can see in [33][34], among others. In this article we present two systems that we consider the most complete in terms of use of framework functions, described as follows.

### *Description*

The Framework was recently tested in the implementation of the two web GIS. The first, the JICA - Japan International Cooperation Agency, together with the Secretary of Planning and Modernization of Public Management of the Government of the State of Tocantins – Seplan—TO in the implementation of the Monitoring System of the Ecological Corridor of the Jalapão Region. The second one was implemented at the Brazilian Institute for the Environment and Renewable Natural Resources – Ibama making a new interface of maps from SisCom -- Shared System available, of which, among others, is responsible for the availability of information in embargoed areas.

Specifically, in the case of the application of the Framework for JICA and the Government of the State of Tocantins, the set of components facilitated the integration of data from different sources. Both institutions dispute spatial data basically in file format, the majority being shapefiles, satellite images in TIF and geographic data in Oracle database with spatial extensions. Nonetheless, the creation of a geographic database with data already validated using PostgreSQL with spatial PostGIS was decided upon.

In this case, the entire application was customized and implemented in one day, having been necessary the creation of the web map services from the distinct sources of data using Geoserver and, later carry out the appointments within the application using the extension files XML. Other files from the same extension were also configured, aiming at altering the logotypes, brand names, title, and subtitle of the application, addresses of services and dynamic captions, among others.

 ISBN: 978-1-61208-251-6 

*Evaluation*

Figures 4 and 5 show two screens taken from the navigation of the application implemented. Figure 4 shows the web GIS with vectoral data only and some of their listed layers dynamically charged by a Geoserver server. The components, "Navigations and Data Tabs", "Layers" and "Menus" are instanced together to the central motor of map navigation.

Figure 5 shows the result of a query in a vectoral database overlayed on the matrix data; they are basically images from the Alos satellite, Avnir sensor and images from the Microsoft Bing catalogue. The "Layers" component was used executing methods such as GetFeatureInfo for retrieving data regarding the layer selected, also it is possible to carry out queries using this same component by way of geographic selections, by simply selecting the area on the screen over which one you wish to research and automatically the parameters are passed on by queries to the interactive map servers.

The result of the queries will always be a grid when the result obtained is greater or equal to two registers. In the Ibama case study, differently from the application implemented for the Jica with geographic extension of just one State, the application used the total extension of the country. The tools for generating graphics were activated and configured aimed at turning the application into a support system for making decisions in terms of the information related to the deforestation and embargoed areas.

Figure 6 shows a screen of the system implemented in the Center for Remote Sensing of the Institute at the moment of consulting georeferenced photography and videos. In this consultation, the components "Navigation and Data Tabs", "Treatment of Geographic DataBase", "Charts", "Layers" and "Media Overlay" are active over the navigation motor.

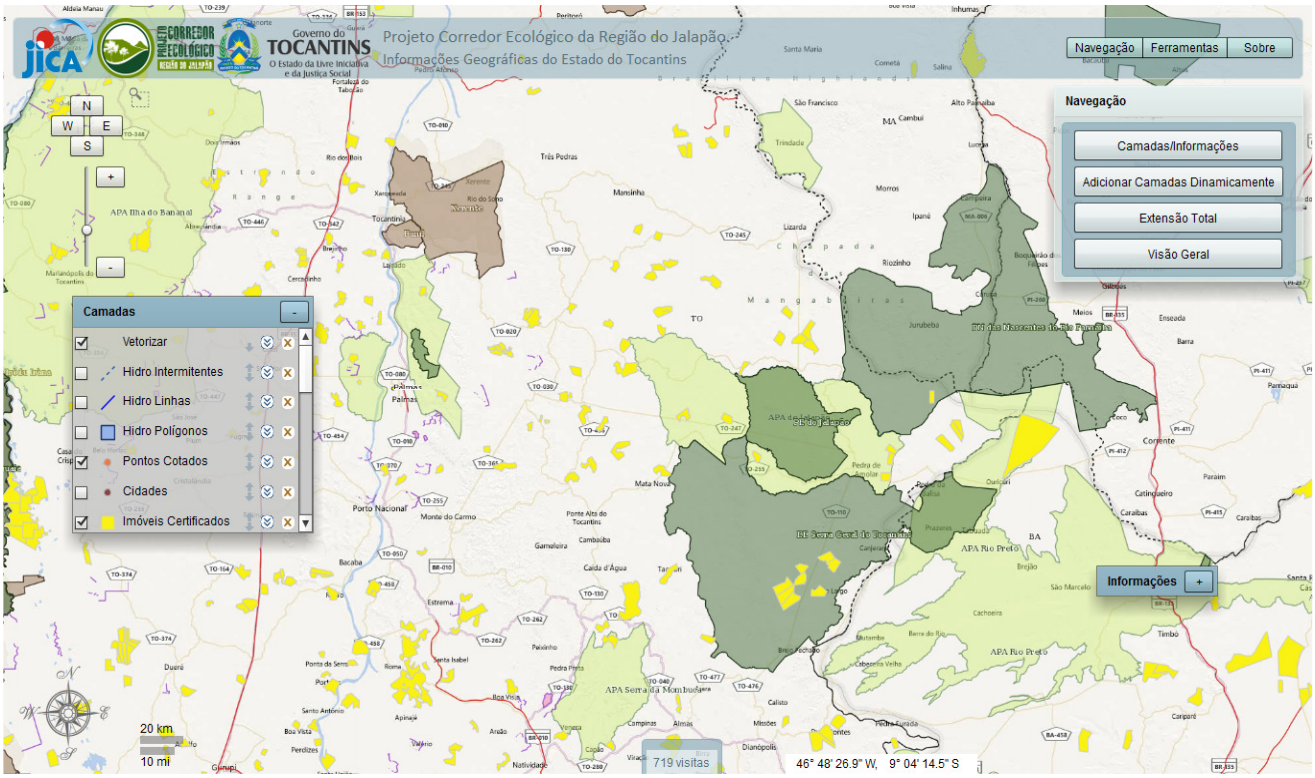



Figure 4 – Basic Navigation.

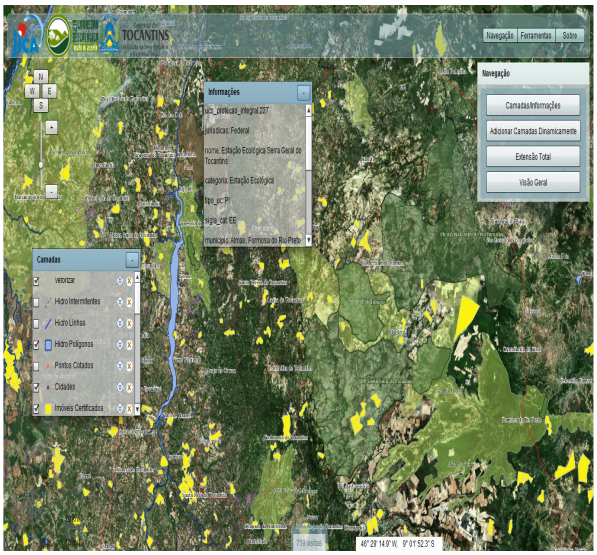



Figura 5 –Raster Navigation and Spatial Query.

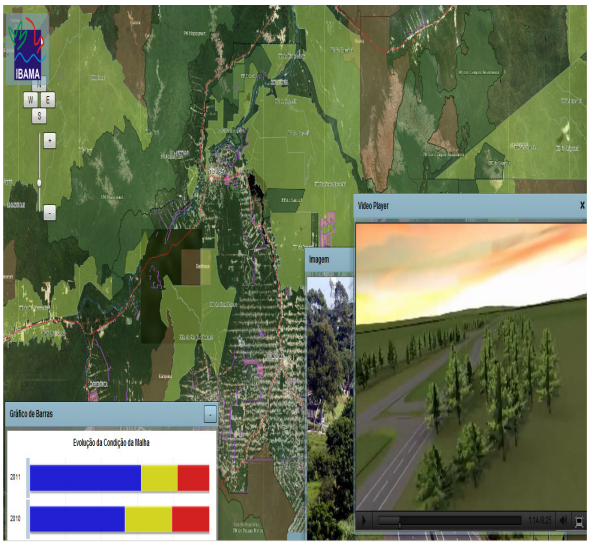



Figura 6 – Ibama Interface with Video Interaction/ Image.

## VI. Conclusion and Future Work

Architecture presented in this paper proposes a tool of rapid implementation and availability of web GIS, with a set of services made available that can easily integrate data of structured systems, such as consumption of information originating from diverse data sources. Textual and vectoral data can be easily manipulated, making not only query possible, but also the insertion, alteration and deletion of information directly in different types of market SGBDs.

The visualization of different media, such as videos, photography, documents and tables originating from map databases was one of the examples of Web 2.0 architecture functioning that impressed users because of its high performance.

The next steps include the development and integration of a 3-D environment, to be written using SDK Flex and OpenGI, to the Framework map environment, as well as, integration with social network tools, also using ActionScript and integration with mosaics and 360º photography and panoramas obtained through Gigapan robots. For the test phase in the 3-D environment data originating from laser profiling will be done through the collection of aerophotogrammatical material in Brasilia in 2010.

## References

[1] Adobe, SDK Flex http://www.adobe.com/devnet/flex/flex-sdk-download-all.html, [accessed 2012-10-20].

[2] J. Arenas and H. Zambrano. Web-based GIS Applications for Government. In Proceedings of the 3rd ICEGOV International Conference on Theory and practice of electronic governance, Bogota, Colombia, 2009.

[3] J. Baumann. Future of Web GIS: An Interview with Pinde Fu. GeoConnection International Magazine, April, 2011.

[4] O. Boucelma and M. Essid. A WFS-Based Mediation System for GIS Interoperability. ACM GIS'02, November 8–9, 2002, McLean, pp. 23-28, 2002.

[5] T. Bressan. Desenvolvimento e integração de um ambiente SIGWEB com ferramentas de software livre. Master Thesis. Federal University of Santa Maria Brazil, 2010.

[6] G. Câmara, L. Vinhas and C. Souza. Free and Open Source GIS: Will There Ever Be a Geo-Linux? Geospatial Free and Open Source Software in the 21st Century. LNGC, pp. 229-245, 2012

[7] S. Dragiéevié and S. Balram. A Web GIS collaborative framework to structure and manage distributed planning processes. Journal of Geographical Systems. Spring-Verlag 6, pp. 133-153, 2004

[8] ESRI, ArcGIS API for Flex http://help.arcgis.com/en/webapi/flex/, [accessed 2012-06-10].

[9] M. Frehner and M. Brandli. Virtual database: Spatial analysis in a Web-based data management system for distributed ecological data. Environmental Modelling & Software 21, pp. 1544-1554, 2006.

[10] P. Fu and J. Sun. Web GIS: Principles and Applications. ESRI Press, 2010

[11] GeoExt. JavaScript Toolkit for Rich Web Mapping Applications. http://geoext.org/, [accessed 2012-10-20].

[12] Google. Maps API for Flash https://developers.google.com/maps/documentation/ [accessed 2012-10-20].

[13] D. Kim and M. Kim. Web GIS Service Component Based On Open Environment. In Proceedings of IGARSS Geoscience and Remote Sensing Symposium. IEEE, pp. 3346-3348, 2002.

[14] A. Longley, F. Goodchild, J. Maguire and J. Rhind. Geographical Information Systems and Science. 2nd Edition. John Wiley & Sons, 2005.

[15] A. Majchrzak and P. More. Emergency! Web. 2.0 the Rescue! . Communications of the ACM, vol. 54, n. 4, pp. 125-132, 2011.

[16] OGC. OpenGIS® Geography Markup Language (GML) Encoding Standard, 2007.

[17] H. Olivera and M. Holanda. A GIS Web with Integration of Sheet and Soil Databases of the Brazilian Cerrado. 2012. In: 7ª Conferência Ibérica de Sistemas e Tecnologias de Informação - CISTI, Madri. Lisboa: AISTI, pp. 735-740, 2012.

 ISBN: 978-1-61208-251-6 

[18] Openscales, http://openscales.org/, [accessed 2012-10-20]..

[19] OSGeo. Open Source Geospatial Foundation. OpenLayers: Free Maps for the Web. http://openlayers.org/, [accessed 2012-10-20].

[20] OSGeo. Open Source Geospatial Foundation. Mapfish. http://mapfish.org/, 2012a

[21] M. Ostergren, J. Hemsley, M. Belarde-Lewis and S. Walker. A Vision for Information Visualization in Information Science. iConference, Seatle, USA, pp. 531-537, 2011.

[22] K. Page, D. Roure and K. Martinez. REST and Linked Data: a match made for domain driven development? In: Proceedings of the Second International Workshop on RESTful WS-REST; Hyderabad, India, pp. 22-25, 2011.

[23] M. Pascual, E. Alves, G. França, H. Roig, T. Almeida, and M. Holanda. An Architecture for Geographic Information Systems on the Web - WebGIS. In: GEOProcessing, The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services, Valencia, pp. 209-214, 2012.

[24] PSPB. Portal do Software Público Brasileiro. I3GEO http://www.softwarepublico.gov.br/ver-comunidade?community_id=1444332, [accessed 2012-10-20].

[25] S. Rao and S. Vinay. Choosing the right GIS framework for an informed Enterprise Web GIS Solution. 13 Annual International Conference and Exhibition on Geospatial Information technology and Applications, India, pp. 19-21, 2010

[26] M. Rhyne. Visualizing Geospatial Data.. In Proceedings of ACM SIGGRAPH GRAPH International Conference on Computer Graphics and Interactive Techniques, 2004

[27] P. Rigaux, M. Scholl and A. Voisard. Spatial Databases with Application to GIS. Elsevier Science, 2002.

[28] H. Shunfu.Web-Based Multimedia GIS for the analysis and visualization of spatial environmental database. In Proceedings of Symposium on Geospatial Theory, Processing and Applications, 2002.

[29] S. Steiniger and A. Hunter. Free and Open Source GIS Software for Building a Spatial Data Infrastructure. Springer, 2012.

[30] R. Wolfgang. Principles and Application of Geographic Information Systems and Internet/Intranet Technology. In Proceedings New Information Processing Techniques for Miitary Systemns. Istanbul, Turkey, 2000.

[31] Vanrijkom http://code.google.com/p/vanrijkom-flashlibs/, [accessed 2012-10-20].

[32] S. Zongyao and X. Yichun. Design of Service-Oriented Architecture for Spatial Data Integration and Its Application in Building Web-based GIS Systems. Geo-spatial Information Science, 2010.

[33] E. Santana, D. Siqueira and M. Holanda. EcoGIS: Sistema de Informação Geográfica para visualização de Dados do Ecossistema do Cerrado Brasileiro. In Proceedings IV GeoNordeste Simpósio Regional de Geoprocessamento e Sensoriamento Remoto, Brazil, pp. 843-848, 2012.

[34] H. Olivera and M. Holanda. A GIS Web with integration of sheet and soil databases of the Brazilian Cerrado. In Proceedings Information Systems and Technologies (CISTI) 7th Iberian Conference on CISTI 2012, Madri, pp. 1-6, 2012.

[35] A. Bozzon, S. Comai and G. Carughi.Conceptual modeling and code generation for rich internet applications. In Proceedings of the 6th International Conference on Web Engineering, pp. 353–360, 2006.

[36] S. William and C. Chadima. ActionScript 3 Design Patterns excerpt: Factory method and MVC. 2007.

 ISBN: 978-1-61208-251-6 

# Urban Perspectives: A Raster-Based Approach to 3D Generalization of Groups of Buildings

Alexey Noskov
Mapping and Geo-Information Engineering
Technion – Israel Institute of Technology
Haifa, Israel
e-mail: mail@alexkarta.com

Yerach Doytsher
Mapping and Geo-Information Engineering
Technion – Israel Institute of Technology
Haifa, Israel
e-mail: doytsher@technion.ac.il

***Abstract*—The suggested method for 3D generalization of groups of buildings is based on rasterization of 2D footprints of the 3D buildings. The rasterization is processed within quarters, which are automatically defined by using Digital Elevation Model (DEM), water objects and roads. The 3D urban perspective is computed based on separate levels of generalization of each quarter as a function of its distance from a pre-defined view point.**

***Keywords-Generalization; 3D urban model; Groups of buildings.***

## I. INTRODUCTION

3D generalization of the urban model is a fast-growing topic. The main types of objects in the 3D city model are buildings. Nowadays, 3D models are used in many disciplines [18]: GPS navigation, desktop and mobile city viewers, geo-simulation, architecture, and many others. The two common problems which usually arise in any discipline are: (1) huge computer resources are required for drawing 3D models based on the original, non-simplified buildings, and (2) 3D models based on the original non-simplified buildings are very detailed and often appear unreadable and overly complex. To resolve both problems we have to generalize the buildings. There are two different tasks in the building generalization process: (1) simplification of a single building, and (2) generalization of groups of buildings. The topic "simplification of a single building" is a widely researched topic [2][9][10][20]; we can describe several different approaches of generalization, all of them valid. In contrast, "generalization of a group of buildings" has been treated, so far, on a very limited level. There are several very close approaches, largely based on the Delaunay Triangulation (DT) [22]. We propose, in concept, another approach for the generalization of groups of buildings, based on rasterization and vectorization operations, which are carried out by sub-dividing the urban neighborhood into quarters.

This paper is structured as follows: the related work is considered in section two, the source data are described in section three, the raster based algorithms of quarter calculations and generalization are considered in sections four and five, the results are evaluated in section sixe, and finally, in the last section the conclusions are detailed.

## II. RELATED WORK

One of the most holistic approaches for the 3D generalization of buildings was described by Xie et al. [22]. The main idea supposes that, within a threshold (distance from a view point), we will generate objects which contain the results of simplification of single buildings, whereas outside of the threshold we will generate objects containing the results of groupings of buildings and simplification of groups of buildings as a single building. An approach of "converting 3D generalization tasks into 2D issues via buildings footprints" was described in [6].

The generalization of 3D building data approach [2], based on scale-space theory from image analysis, allows simplifying all orthogonal building structures in one single process. Another approach [20] considers buildings in terms of Constructive Solid Geometry (CSG). In [22], an approach was proposed which realized 3D single building simplification in 5 consecutive steps: building footprint correction, special structure removal, roof simplification, oblique facade rectification and facade shifting. A very interesting approach was proposed in [9] and [10]. In this approach, geometric simplification was realized by remodeling the object by means of a process similar to half-space modeling. Approximating planes are determined from the polygonal faces of the original model, which are then used as space dividing primitives to create facade and roof structures of simpler shapes.

The second aspect of 3D generalization of an urban environment is the generalization of groups of buildings. 3D generalization of groups of buildings is mentioned in several publications (e.g., [3][5][6][21]). These papers describe different approaches of 3D grouping and group generalization: grouping of building models (using the infrastructure network) and replacing them with cell blocks, while preserving local landmarks [3]; "express different aspects of the aggregation of building models in the form of Mixed Integer Programming problems" [5]; and, grouping of building models "with a minor height difference and the other with a major height difference" [6].

2D building generalization algorithms should also be considered as being used by researchers for a 3D building group generalization. A holistic and automated generalization method based on a pseudo-physical model was considered in [8]. An approach based on Delaunay triangulation, Graph and Gestalt theory was described by Li et al. [16].In the above-mentioned publications, different approaches were considered, but we can identify some

 ISBN: 978-1-61208-251-6 

common ideas which are important for most research in this area.

In most cases it is very useful to generate levels-of-detail (LOD); normally, researchers use 3 or 4 LODs ([1], [17] and [20]). LODs are widely used in 3D video games, usually for detailed objects; more simplified objects are created for saving processor load and virtual memory [17]. Usually a detailed object has references to several simplified versions (at different levels of simplification), so that if the object stays near the view point, the most detailed version of the object is used, and as the object is located further from the view point the more simplified object is used.

It is very popular to use CityGML standard for 3D urban models ([4], [7], [11], [12], [13], [14], [15] and [19]). This format supports many useful possibilities, which are very important for working with 3D urban models (e.g., LODs, topology, semantics etc.).

## III. SOURCE DATA

For implementing and testing our approach, the free geodata of the city of Trento, Italy was used. The buildings (with individual heights), water objects and land relief (DEM) were downloaded from the website of Trento Municipality [23]; roads were downloaded from the OpenStreetMap website [24]. On the map of Trento (see Figure 1) the buildings are depicted as gray areas; the extent of the maps in Figure. 2, Figure 3 and Figure 4 are marked with a blue square; the extent of the map on Figure 8 as a red square, and the view point and view direction of Figure 9 as an orange circle and arrow.

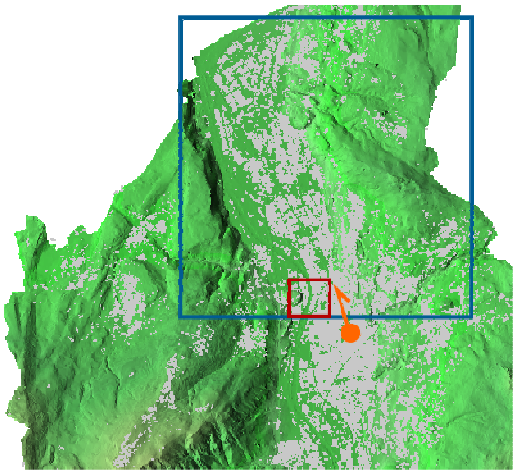


Figure 1. Map of Trento.

## IV. CALCULATION OF QUARTERS

Finding a realistic method of simplification is a very important issue in generalization. One of the more common problems is when buildings are being joined through obstacles such as wide roads or rivers. In this case, buildings do not have to be joined to each other, and these buildings from the two sides of the obstacle, should be merged with other, more distant objects, which are, however, located on the correct side of the obstacle. To resolve this problem, we decided to split the urban space into quarters which are divided by the main, significant objects. These objects cannot be involved in the generalization itself.

To calculate quarters, we decided to use the slope of the terrain, water objects and roads. In Trento, it was found that buildings are positioned only on areas with a slope smaller than 30 degrees. Accordingly, areas with slopes greater than 30 degrees of the terrain were excluded. The main road types were used as dividers (excepting negligible roads such as 'footway', 'pedestrian', 'service' and others). The third class of objects for defining the quarters was the water objects. All these three classes – slopes, roads and water objects - were merged into one raster map with 1 meter resolution (which has been found to be adequate for small scale urban generalization). For line objects (roads, narrow rivers) we used a minimal width for the objects which is equal to 1 meter (1 pixel).

The raster map of the merged objects is the base for quarter calculating; further processing can be divided into several consequent steps.

The raster transformations for splitting the city into quarters have been selected because the standard vector approaches (e.g., polygons based on vector roads) have several limitations. The source vector road data may contain features such as unfinished roads, dead end roads, etc., features affecting its topological correctness. Splitting the area into quarters based on these data might result in a very complex polygonal map, containing artifacts. In contrast, the raster transformations approach enables to exclude most of the artifacts and the unnecessary bounds and vertices. As the width of the narrowest roads is about 2-3 meters, the resolution of the raster maps has been defined as 1 meter. Accordingly, the quarter map is composed of polygon bounds which coincide approximately with the road centerlines (±1 meter), as well as not intersecting the buildings.

### A. *Region growing of base features*

All pixels of the merged objects got the value “1”; empty space on the raster map got the “Nil” value. Each group of pixels with value “1” has been expanded by adding one pixel (1 meter) and the results are depicted in Figure 2.

### B. *Inverting of pixel map*

At this step, the values of pixels were inverted (“1” to “Nil” and vice versa), which results with many pixel areas with the value “1” which are split by “Nil” pixels.

### C. *Defining quarter areas having unique values*

To set a unique pixel value to each quarter area, we vectorized the raster map. Each vector that defines a polygonal object got a unique integer identifier. Polygonal objects with areas of less than a pre-defined threshold were removed (in our case, an area of a threshold of 600 square meters was found to give adequate results). Then the vector map was rasterized. For raster values, polygon identifiers were used. As a result (see Figure 3), we got a raster map with group of pixels (a "quarter") and each group (which is

 ISBN: 978-1-61208-251-6 

separated by "Nil" pixels from the adjacent group) got a unique integer identifier.

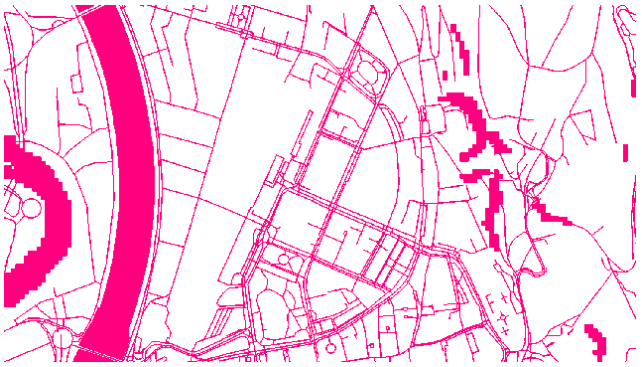


Figure 2. Non-Nil Pixel Groups which Split the City Space into Quarters.

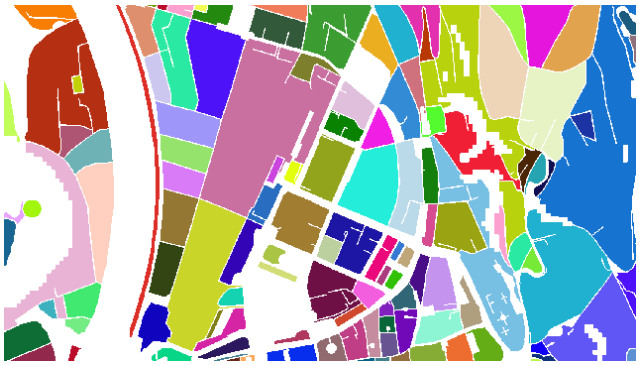


Figure 3. Inverted Raster Map with Unique Pixel Values.

### D. *Region growing raster map and final vectorization*

At this stage, the raster map has a lot of empty areas: quarters contain empty areas, and spaces between quarters are empty. The quarters look too complex and contain too many artifacts. To resolve this problem, the non-Nil pixel groups were expanded. Consecutive pixels were added to each unique pixel area until a non-Nil neighbor pixel or threshold is achieved (in our case, 15 pixels/meters out of the current pixel group). Then the raster map was vectorized.

As a result (see Figure 4), we got the final raster map which contains continuous groups of pixels, where each group has a unique integer identifier. Then, the raster map is vectorized into a topologically correct quarter polygons with minimum artifacts. The total number of quarters in Trento was 1,431.

## V. GENERALIZATION

The fact that, in urban areas, most (if not all) of the buildings have orthogonal sides, is the background of our raster-based generalization approach. Usually, in adjacent areas (quarters in our case), buildings would be spatially oriented in the same direction. Therefore, the generalization process consists of defining the typical azimuth of buildings' sides for each quarter. Once a typical azimuth is known, by applying the rasterization process in this direction, the staircase-type appearance of lines, or legs of closed polygons, which is very common in the rasterization processes, can be eliminated. A non-rotated rasterization (parallel to the grid axes) while the buildings are positioned in another orientation will result in a staircase-type appearance of the bordering lines of the buildings and too many unnecessary vertices which will prevent achieving a smooth geometry of the generalized objects.

Figure 4. Final Quarters.

### A. *Defining the azimuth of buildings' sides*

In urban areas, most of the buildings have orthogonal sides; thus, it is possible to define the average spatial orientation of the buildings. Within each quarter, the azimuths of all the buildings' sides were computed. For each building in the quarter, the longest side and its azimuth were identified. Then, all the azimuths of the other sides are rotated by 90 degrees (clockwise) again and again; and the rotated azimuths (and their lengths) were put in one list. The list was sorted by lengths, and then lengths with the same azimuths (up to a predefined threshold) were averaged. A threshold of 1 degree when looking for close buildings' side azimuths has been found to give satisfactory results. A weighted average of the azimuths of the longest lengths of all the buildings within a quarter is used to define the general orientation of all the buildings of the quarter.

### B. *Rotation and rasterization of the buildings in a quarter*

As mentioned above, and in order to significantly reduce the number of vertices of the generalized building and achieve a more realistic appearance of these simplified objects, the rasterization should be carried out in the spatial orientation of the buildings. A rasterization which is spatially oriented parallel to the grid axes will define the buildings which are not oriented parallel to the grid axes in a staircase-type appearance of the buildings' sides. Accordingly, all the buildings within a quarter were rotated counter-clockwise at the angle of the general orientation of all the buildings of the quarter. Then, the rotated buildings were rasterized using a certain pixel size resolution (as explained in the next section). Each pixel with more than half its area covered by the original buildings gets the value "1"; otherwise it gets the value "Nil". Figure 5 shows the result of this stage.

The level of the generalization is a function of the pixel size rasterization process - the greater the pixel size, the greater the degree of generalization. Accordingly, each

 ISBN: 978-1-61208-251-6 

quarter has been generalized at several levels of rasterization, resulting in several layers of different levels (level-of-detail) of generalized buildings for each quarter. Based on the original data of Trento, and according to our analyses, we found that using pixel size resolutions of 10, 15, 20, 25, 30, 40, 50 and 60 meters produces satisfactory results of a continuous and consecutive appearance of the level-of-detail of the generalized buildings.

Figure 5. The Generalization Process of Buildings in a Quarter: Original Buildings (left); Rotated Quarter and the Generalized 10 meter Rasterized Buildings in red (middle); Final Result (right).

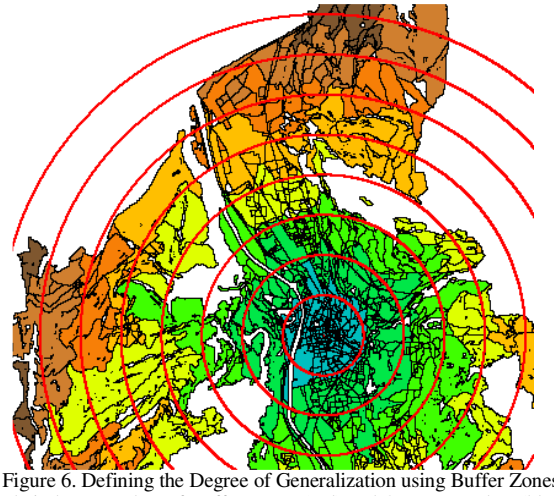


Figure 6. Defining the Degree of Generalization using Buffer Zones: Red circles – Borders of Buffer Zones (red) and Quarter Borders (black).

To draw a 3D perspective of the city with the generalized buildings, the position of a view point has to be defined. Then we built buffer zones around the view point. The buffer zones defined the distances (practically, range of distances) from the view point to each quarter. Then, we set the degree (resolution) of generalization of the buildings for each quarter. Figure 6 depicts the degree of generalization for each quarter, where the colors indicate the degree of the generalization. The relationship between the distances from view point, pixel size generalization, and the colors, are described in Table I. Finally, we merged all the separate generalized layers of all the quarters into one map (see Figure 8) for further 3D visualization. The division of distances from the view point into a scale of continuous intervals was based on several tests, which enabled us to draw a realistic and continual 3D model or perspectives. The results of a 3D visualization, and comparison of the 3D perspectives with the original buildings and with the generalized buildings, are presented in "Figure 9".

TABLE I. DISTANCES FROM THE VIEW POINT, RESOLUTIONS OF GENERALIZATION, AND COLORS.

| Distances from view point, meters | Resolutions of generalization, meters | Background colors of the map in "Figure 6" |
|---|---|---|
| 0 - 1000 | original buildings | |
| 1000 - 2000 | 10 | |
| 2000 - 3000 | 15 | |
| 3000 - 4000 | 20 | |
| 4000 - 5000 | 25 | |
| 5000 - 6000 | 30 | |
| 6000 - 7000 | 40 | |
| 7000 - 8000 | 50 | |
| >8000 | 60 | |

## VI. NUMERICAL EVALUATION

Table II presents the number of geometry primitives and the speed of the visualization process as a comparison between the original data and generalized data. As we can see, there is a significant reduction in visualization speed and in the number of polygons and nodes.

TABLE II. RESULTS OF THE GENERALIZATION.

| Parameter | Original building layer | Generalized building layer used for 3D visualization |
|---|---|---|
| Number of nodes | 114,648 | 34,391 |
| Number of polygons | 46,339 | 14,956 |
| Speed of 3D visualization, second | 6.6 | 1.2 |

To evaluate the quality of generalization the mean coefficient of building compactness was calculated for each resolution of generalization (see Figure 7). The coefficient of compactness of a single building is equal to $\alpha=P2/(4*\pi*A)$, where P – perimeter, A – area ($\alpha=1$ for a circle, $\alpha=1.27$ for a square).

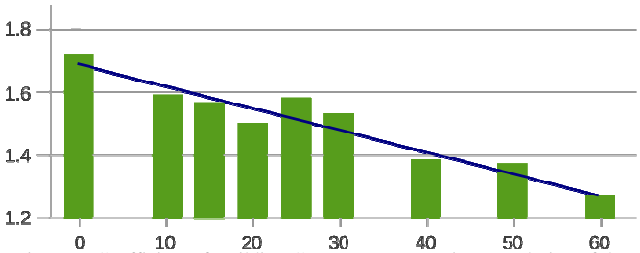



Figure 7. Coefficient of Building Compactness: X-axis – Resolution of the Generalization, 0 – Original Buildings.

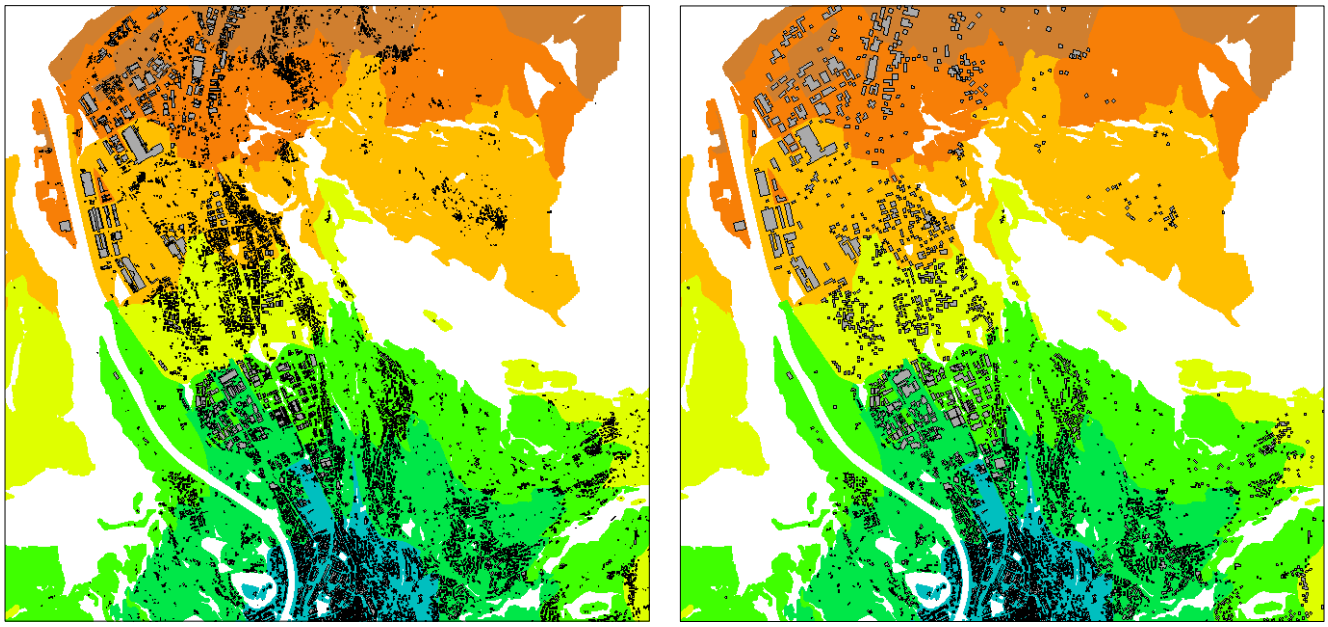


Figure 8. The Northern half of Trento with the Original Buildings (left) and with the Generalized Buildings (right): Different Levels of Generalization and background Colors are according to Table I.

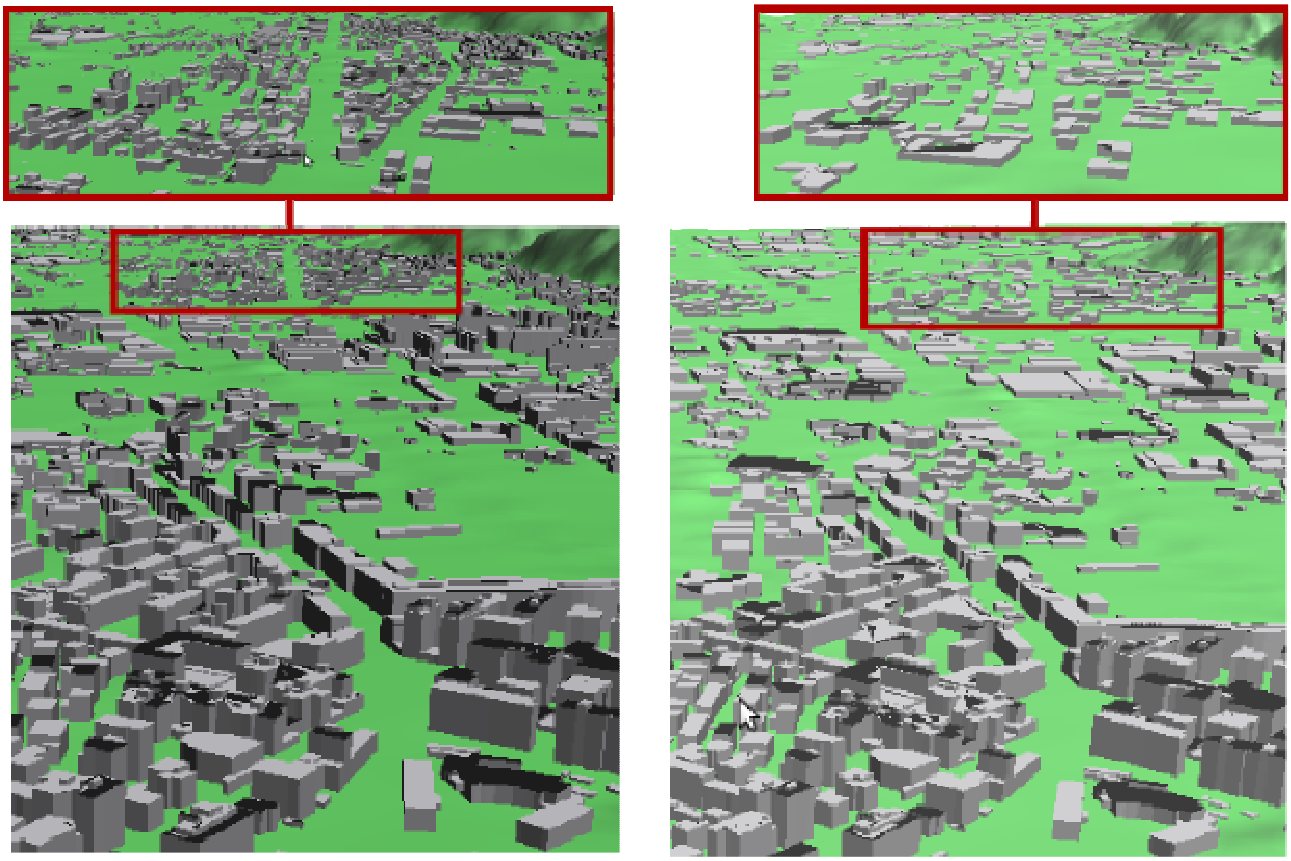


Figure 9. 3D Perspective with the Original Buildings (left) and with the Generalized Buildings (right). Zoomed Areas are Marked in Red.

 ISBN: 978-1-61208-251-6 

In Figure 7, we can see that the coefficients of the buildings' compactness decreases significantly from 1.71 to 1.27, which demonstrates the efficiency of the approach.

The method and the process have been developed by using a standard PC (DELL Vostro 3550), 4 processors: Intel® Core™ i3-2310M CPU @ 2.10GHz, with 1.8 GB Memory. In addition, Ubuntu operating system, GRASS GIS, Bash and Python programming languages have been used.

## VII. CONCLUSION AND FUTURE WORK

The raster-based approach of the method is based on standard tools of rasterization, vectorization, region growing, and overlaying. The main advantage of the developed method is the ability to simplistically and efficiently generalize buildings at different levels, achieving variable, but continuous, level-of-detail of the buildings as a function of the depth of the plotted perspectives. The continuity of the generalized product is achieved by subdividing the area of the city into quarters, which take into account the significant objects affecting the process. As a result, the generalized 3D model does not contain unreadable and overly detailed separate buildings on the one hand, and is able to merge further groups of buildings on the other. At the same time, even though the buildings are simplified, the model maintains the geographical correctness and specifications of the urban area. The developed method helps reduce the time, and the required computer resources, for drawing 3D models or perspectives of a city or urban areas.

The current solution is based on a rigid subdivision of the processed area into quarters. This approach of rigid quarters limits the maximum level of generalization to the minimal quarter size on the one hand, and is unable to take into account the density and distribution of the buildings during the process on the other hand. Further research will be focused on improving the suggested model toward a more dynamic and flexible solution. In addition, it is also planned to improve the mechanism of defining the heights of the generalized objects. While in the suggested approach we set a single height to a group of neighbor pixels, it will be more precise to define an individual height to each pixel and only then classify the pixels into groups, which will enable us to achieve a more realistic result.

## REFERENCES


[1] J. Döllner and H. Buchholz, "Continuous Level-of-detail Modeling of Buildings in 3D City Models," in GIS'05 Proceedings of the 13th Annual ACM International Workshop on Geographic Information Systems, Bremen, Germany, 2005, pp. 173-181.

[2] A. Forberg, "Generalization of 3D Building Data Based on a Scale-Space Approach," ISPRS Journal of Photogrammetry & Remote Sensing, vol. 62, 2007, pp. 104-111.

[3] T. Glander and J. Döllner, "Abstract Representations for Interactive Visualization of Virtual 3D City Models," Computers, Environment and Urban Systems, vol. 33, September 2009, pp. 375-387.

[4] G. Gröger, T. H. Kolbe, and L. Plümer, "City Geographic Markup Language," Approved Discussion Paper of the Open Geospatial Consortium, 2006.

[5] R. Guercke, T. Götzelmann, C. Brenner, and M. Sester, "Aggregation of LoD 1 Building Models as an Optimization Problem," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, 2011, pp. 209–222.

[6] S. He, G. Moreau, and J. Martin, "Footprint-Based 3D Generalization of Building Groups for Virtual City," GEOProcessing 2012: The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services, Valencia, Spain, 2012, pp. 177-182.

[7] U. Isikdag and S. Zlatanova, "Towards Defining a Framework for Automatic Generation of Buildings in CityGML Using Building Information Models," 3D Geo-Information Sciences Lecture Notes in Geoinformation and Cartography, Part II, 2009, pp. 79-96, DOI: 10.1007/978-3-540-87395-2_6.

[8] J. Joubran Abu Daoud and Y. Doytsher, "An Automated Cartographic Generalization Process: A Pseudo-Physical Model," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVII, part B2, Beijing, 2008, pp. 419-424.

[9] M. Kada, "3D Building Generalization Based on Half-Space Modeling," Joint ISPRS Workshop on Multiple Representation, Hannover, Germany, 2006, pp. 58-64.

[10] M. Kada, "Automatic Generalisation of 3D Building Models," in Proceedings of the Joint International Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada, 2002, pp. 32-38.

[11] T. Kolbe, "Representing and Exchanging 3D City Models with CityGML," in Proceedings of the 3rd International Workshop of 3D Geo-information, Seoul, Korea, 2009, pp. 15-30.

[12] T. Kolbe, C. Nagel, and A. Stadler, "CityGML—OGC Standard for Photogrammetry?," In: Fritsch, D. (Ed.), Photogrammetric Week 2009, Wichmann Verlag, Heidel-berg, pp. 265–277.

[13] Kolbe T.H. and Gröger G., "Towards Unified 3D City Models," in Schiewe, J., Hahn, M, Madden, M, Sester, M. (Eds.): Challenges in Geospatial Analysis, Integration and Visualization II. Proceedings of Joint ISPRS Workshop, Stuttgart, Germany, 2003, pp. 41-49.

[14] T. H. Kolbe and G. Gröger, "Unified Representation of 3D City Models," Geoinformation Science Journal, vol. 4(1), 2004.

[15] T. H. Kolbe, G. Gröger, and K. Plümer, "CityGML – Interoperable Access to 3D City Models," in Proceedings of the First International Symposium on Fachbeiträge Geo-information for Disaster Management, Delft, Netherlands, 2005, pp 883-899.

[16] Z. Li, H. Yan, T. Ai and J. Chen "Automated Building Generalization based on Urban Morphology and Gestalt Theory," Int. J. Geographical Information Science, vol. 18, 2004, pp. 513-534.

[17] D. Luebke, M. Reddy, J. D. Cohen, A. Varshney, B. Watson, and R. Huebner, "Level of Detail for 3D Graphics," The Morgan Kaufmann Series in Computer Graphics, Edition 1, 2002.

[18] N. Shiode, "3D Urban Models: Recent Developments in the Digital Modelling of Urban Environments in Three-Dimensions," GeoJournal, vol. 52(3), 2000, pp. 263-269.

[19] A. Stadler, C. Nagel, G. König, and T. H. Kolbe, "Making Interoperability Persistent: A 3D Geo Database Based on CityGML," 3D Geo-Information Sciences. Lecture Notes in Geoinformation and Cartography, Part II, 2009, pp. 175-192.

[20] F. Thiemann, "Generalization of 3D Building Data," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (Part 4), 2002.

[21] M. Trapp, T. Glander, and H. Buchholz, "3D Generalization Lenses for Interactive Focus + Context Visualization of Virtual City Models," in Proceedings of the 12th International Conference Information Visualization, London, 2008, pp. 356 - 361.

[22] J. Xie, L. Zhang, and J. Li, "Automatic Simplification and Visualization of 3D Urban Building Models," International Journal of Applied Earth Observation and Geionformation, vol. 18, 2012, pp. 222–231.

[23] http://webapps.comune.trento.it/ambiente/ [retrieved: 12, 2012].

[24] http://www.openstreetmap.org/ [retrieved: 12, 2012].


 ISBN: 978-1-61208-251-6 

# Simple Methods for Reasoning about Behavior Patterns on Graphs Given Extremely Sparse Observations

R. Paul Wiegand
*Institute for Simulation & Training*
*University of Central Florida*
*Orlando, FL, USA*
*Email: wiegand@ist.ucf.edu*

Steven D. Prager
*Department of Geography*
*University of Wyoming*
*Laramie, WY, USA*
*Email: sdprager@uwyo.edu*

***Abstract*—We consider the situation where fixed observations of moving entities are sparse and the goal is to learn as much as possible about their *patterns of activity*, before and after such observations (e.g, cameras at a few intersections in a city). Here we present a method for estimating probable paths within a network given a limited set of vertex observations and limited *a priori* assumptions about individual entity behavior. We divide the process of analysis into two phases: a *learning phase* in which aggregate information about many entities is obtained and used to construct simple models given potential observations, and a *reasoning phase* in which resampling methods produce probable paths a specific entity may have taken. To accomplish this, we extend a fair and efficient method for randomly selecting unconditioned paths within a network in order to draw paths conditioned on limited, partial observations. The methods are validated by analyzing hypothetical observations of entities moving on an existing city street network. Our results show the scaling properties of this approach by optimizing the locations of different numbers of fixed potential observation points to obtain as maximal coverage of the area as possible. We then construct a variety of models based on an extremely sparse observational scenario and demonstrate quantitatively and visually that these simple methods, combined with structural information inherent in the graph itself, can provide a great deal of context information about an individual entity's possible movement patterns.**

***Keywords*-sparse observations; fair paths; path analysis**

## I. INTRODUCTION

Many GIS-based methods for analyzing the behavior patterns of entities (e.g., people, cars, etc.) on both traditional and complex functional networks incorporate extensive information about individual entity movement within the network, often assuming one has access to GPS tracking data or other such detailed information. Frequently, the goal in such cases is to take existing data and use it to predict *future activity* [1], [2].

Instead, we consider the situation where observations are sparse and the goal is to try to learn as much as possible about a *pattern of activity* before and after such observations. For example, suppose there are fixed CCTV camera locations at a few intersections, scattered throughout a city. Given some highly limited general knowledge of potential paths taken (e.g., a distribution of typical path lengths) and some specific observations of a particular entity at two or three of the intersections via the cameras, what are the most probable paths the entity might have taken?

In this paper, we construct and employ a simple method for drawing probable paths within a network given a handful of vertex observations and limited *a priori* assumptions about individual entity behavior. Our approach is to breakdown the process of analysis into two phases: a *learning phase* in which aggregate information about many entities is obtained and used to estimate path length distributions conditioned on potential observations, and a *reasoning phase* in which resampling methods are used to combine partial observations about a specific entity with global knowledge to provide an analyst probable paths an entity may have taken. Though it is natural to think about these path-patterns as movement in a physical network such as city streets, our method makes no such assumptions.

This paper is developed around a fair and efficient (polynomial) method for randomly selecting unconditioned paths within a network in a probabilistically correct way. Our algorithm (DRAWPATH) is constructed using elementary concepts in graph theory and probability. We extend this model to draw paths conditioned such that only paths that contain specific observations, in a particular order (though not necessarily consecutive) will be drawn. We validate our approach by considering the analysis of hypothetical movement patterns of entities operating on the city street network in the western half of Laramie, Wyoming.

To determine the best coverage of the observations in the town, we employ an evolutionary algorithm (EA) [3] to optimize the locations of a limited number of fixed potential observation points. The objective function used by our EA relies on the same simple graph-theoretic notions on which our DRAWPATH algorithm is based. We have a simple way to estimate the probability that a path from the general, aggregate movement distribution will pass through at least one of the observations. We show that the coverage we obtain scales well even as the number of potential observation locations is greatly reduced.

 ISBN: 978-1-61208-251-6 

We then focus on an extreme case where there are only two fixed positions in the town that we can use to analyze movement. Several models are constructed based on these two fixed positions, and these to demonstrate quantitatively and visually that a great deal of context information about an entity's possible movement patterns can be obtained in spite of the sparsity of observations. This provides preliminary evidence for our position that the graph itself contains a great deal of information, and that given a simple aggregate movement model and a few individual entity specific observations, it is still possible to obtain a strong understanding of probable movement patterns for the entity by exploiting structural information within the graph itself.

This paper is organized as follows. The next two sections provide context by discussing related literature and problem formalization, respectively. Section IV explains are simple, traditional graph theory based methods for drawing distributionally correct fair random paths. After this, we indicate how much coverage is possible with few observations by optimizing potential observation locations for maximal coverage, then in Section VI we discuss how prior distributions for potential movement patterns through sparse observations can be learned. Finally, we explain how one can put all this together to reason over potential paths given sparse observations and provide some concluding remarks.

## II. Background and Motivation

Understanding entity movement in space and time builds on early work in regional geography regarding the concept of space-time paths [4]. The space-time path — a "trace" of an individual entity's movement in space, through time — is comprised of the set of ordered (in time) observations where entity locations are recorded and the corresponding path segments between those locations are inferred [5]. Where space-time path approaches are limited, however, is when the set of observations are limited and serve to capture only a portion of a path.

Given the limitations of space-time path approaches when much of the requisite entity movement data are unobserved, alternative mechanisms to understand individual movement are required. As technology and computational approaches have advanced, focus has shifted toward modeling movement in space using more bottom-up, agent-based approaches [6]. For example, individual pedestrian movements have been modeled at the "microscopic" scale [7] in order to examine the role of fundamental behaviors in governing movement characteristics. Other agent-based approaches have emphasized crowd level behaviors [8], in order to contextualize movement given larger, agglomerative behaviors. These bottom-up approaches, however, often assume little knowledge regarding the geographic space or network neighborhood in which the entities are operating.

Understanding the space in which entities are moving is an important aspect to understanding their potential behavior. Space-syntax studies demonstrate that network and related areal topologies have demonstrable effect in how space influences movement [9]. The environment, whether unconstrained (e.g., an isotropic surface), constrained (e.g., a network), static, or dynamic, is an important determinant in the driving the modeling approach [10]. Recognition that certain movement patterns or *typologies* emerge as a function of the space in which movement is occurring provides a basis for understanding movement from limited observations [11].

Using limited observations to best cover a given space (network or otherwise) and understand movement occurring therein is thus an important aspect of this work. As with the space-time path, early work in examining maximum covering problems in a geographic context emerged in a formalized way in regional geography studies [12]. Maximal covering location theory has informed work ranging from identification of best locations for security cameras in both static [13] and dynamic environments [14].

In contrast to the high observation model of sensor network approaches for understanding space-time paths, low observation models require as much information as possible that can be exploited from other sources. In the following sections we formalize our approach for leveraging information regarding the network space in which movement is occurring and the use of a simple movement model of describing individual movement within the given environment.

## III. Formalizing the Problem

As described in the introduction, the fundamental challenge we are considering is to determine the probability that some specific person (for example) has traversed some path given a graph, some observations of the person on that graph, and a some kind of more general underlying distribution regarding path properties. Let us clarify this problem both by providing a formal description of the problem and by example. For descriptive purposes, our examples refer to some particular path taken by an individual named *Bill*.

Let $G = (V, E)$ be a digraph with vertices $V$ and edges $E$. Graphs are given in the problem description, and they represent some kind of network-based structural information available in the problem itself. E.g., it might encode topological information about Bill's home town street network.

We write a path simply as an ordered set of vertices, $p = \langle v_0, v_1, \ldots, v_k \rangle$ such that $\forall v_i \in p, v_i \in V$ and $\forall v_i, v_{i+1} \in p, (v_i, v_{i+1}) \in E$. That is, a path is a sequence of connected vertices. Note that paths *can* contain the same vertex multiple times (i.e., the path can cycle). We define the *length* of a path as the number of vertices in the path[1]. A path represents every graph point on some route that a person traverses within the graph. For example, Bill leaves the bank, moves south a block to the corner gas station, then

[1]Technically, in graph theory, path length is traditionally defined as the number of hops taken, $k-1$ where $k$ is the number of vertices in the path; however, we keep everything vertex-centric for consistency.

 ISBN: 978-1-61208-251-6 

heads east to the intersection of main and $1^{st}$, then moves east once more to his apartment. His path's length is three.

In addition, we have the concept of an *observation*, a fixed vertex at which we can obtain information about a specific person on a specific path. We may have multiple observations regarding the same path. While we retain *order* information about observations, one simplification we make is to ignore time. We distinguish between two types of observations, *positive observations* and *negative observations*. Positive observations are an ordered collection of sightings of an individual at specific vertices for the path, $\langle o_1, o_2, \ldots, o_m \rangle$, where there exists some path $p$ such that for all $o_i$ in that collection, $o_i \in p$. The set of negative observations is an unordered collection of vertices through which an individual *did not* pass. For simplicity, we use $M$ to notate the complete collection of observations. As an example, consider a case where we observed Bill at the gas station and later at the intersection of main and $1^{st}$ avenue, though we know he was not seen by the camera at the bank. Most of his path, including his start and end point, are unknown to us; however, we have two positive observations and one negative observation regarding his path.

In addition to representing individual behavior, we are interested in collecting aggregate information about possible paths. Currently, we simplify things and consider only path length for the prior distribution of path properties, and we assume that (unconditioned) path lengths are Poisson distributed. That is, for some path length $k$:

$$\Pr\{k\} := \frac{\lambda^{(k-1)} e^{-(k-1)}}{(k-1)!}, \quad k > 0$$

The $k-1$ adjustment is due to our having defined the path length as the number of vertices. Conceptually, the number of null-paths, paths that contain *no* vertices at all, is not a sensible number to consider. For convenience, we say that $\Pr\{k = 0\} = 0$. Under the optimistic conditions provided by the problem definition, namely that a set of historical path lengths are provided, we can determine $\Pr\{k\}$ by simply computing the expected length parameter $\lambda \approx \frac{1}{|K|} \sum_{i=1}^{|K|} k_i$. For our experiments, we assume this is done and specify $\lambda$ directly. In this paper, we use $\lambda = 6$ for all experiments.

We refer to $\Pr\{k\}$ as the *prior path length distribution* and $\Pr\{k|M\}$, the probability of a path of length $k$ if we make observations $M$, as the *posterior path length distribution*. We break the problem into two phases, *distribution learning* and *path resampling*, and define them separately. These are described in Section VI and VII, respectively.

PROBLEM *1:* **Distribution Learning**

**Given:** a graph $G$, hypothetical observations $M$, and a set of historical path lengths, $K$ (where $\forall k_i \in K, k_i \in \mathbb{N}^+$).

**Find:** prior path length distribution $\Pr\{k\}$ and posterior path length distribution $\Pr\{k|M\}$.

PROBLEM *2:* **Reasoning via Path Resampling**

**Given:** a graph $G$, real observations $M$, and the model distribution $\Pr\{k|M\}$

**Find:** a set of paths $P$ such that $\forall p_i \in P, \Pr\{p_i|M\}$ is "sufficiently large".

## IV. FAIR PATH SAMPLING WITH BASIC GRAPH THEORY

Ultimately, our efforts require efficient estimation methods for drawing paths that take graph structure into consideration. One might simply produce random paths via a random walk on the graph. For example, a starting node might be selected at random, a random neighbor might be selected from that vertex, repeating this process until a designated criteria has been satisfied (e.g., a path length drawn *a priori* according to some known distribution). However, ensuring that constraints on the graph do not unintentionally bias such walk is non-trivial, particularly when the graph is directed. Alternatively, one might simply enumerate all possible distinct paths (addressing loops in some well-defined way) and select a path from this set. Unfortunately, there may be a combinatorially large number of potential paths. Moreover, this set may be quite sparse with respect paths that contain the observations.

Instead of these approaches, we turn to traditional graph theory. Note that connectivity in a typical graph can be represented using an adjacency matrix, $L$, containing binary values, where a 1 in some cell $L_{uv}$ indicates that there is a directed edge in the graph from vertex $u$ to vertex $v$, and a 0 indicates that there is no such edge. The matrix $L$ can also be interpreted as indicating all simple paths of length precisely 1 between any pair of vertices. We can apply matrix multiplication to produce a new matrix of the same dimensions, $L^2$, which then gives us the number of paths of length 2 connecting each pair of vertices. In general, $L^k_{uv}$ gives the number of distinct paths of length $k$ that connect $u \rightsquigarrow v$ [15]. Our methods use this property in a variety of ways. We refer to this as a "*path-count*" calculation.

We use this path-count calculation to estimate the probability that a path drawn according to our aggregate model distribution (in this case, a path in the graph whose length is Poisson distributed) passes through at least one of some set of observations. For a given length, $\ell$, the total number of paths of that length can be computed by summing all values in the $L^\ell$ matrix for a particular graph. Paths that *do not* pass through the one of the vertex observation points cannot traverse an edge to or from any of the potential observation points, so we can produce a new graph by removing all the potential observation point vertices and their out-going and in-coming edges. The total number of paths for a given length for this graph can be computed in the same way as for the original graph. The difference between these is the total number of paths of length $\ell$ that passed through at least one of these edges, and the ratio of this over all possible $\ell$-length

 ISBN: 978-1-61208-251-6

paths in the original graph is the probability of drawing such a path given the length, $\Pr\{D|k=\ell\}$, where $D$ is the event of drawing a path going through at least one of some set of observations. Thus $\sum_{i=1}^{\infty}\Pr\{D|k=i\}\Pr\{k=i;\lambda\}$ yields the probability of a Poisson distributed path with expected length $\lambda$ going through at least one of the observations. We estimate this by summing such values until some sufficiently large path length or until the $\Pr\{k=i;\lambda\}$ is very small.

Secondly, the path-count metric provides a means of "fair" sampling for paths given a distribution of path lengths. Algorithm 1 presents the pseudocode for DRAWPATH.

**Algorithm 1** DRAWPATH (graph connection matrix $L$, path length distribution)

```
Randomly select path length ℓ according to distribution
Compute the L^ℓ path-count matrix
Randomly select a source-destination pair for path p = ⟨s, ⇝, t⟩
    proportionate to the L^ℓ path counts
Let a = s
Let b = t
for i = ℓ − 1 down to i = 2 do
    Compute the L^i path count matrix
    Form p_a, the 1-step ab intersection of this and L as follows:
        Let r_a be the a^th row of L^i
        Let r_b be the b^th column of L^i
        Let c_b be the b^th column of the L connectivity matrix
        Let p_a = r_a × c_b, using element-by-element multiplication
    Randomly select vertex, x, proportionate to p_a values
    Insert x following a in p and let a = x
end for
```

A process following such an algorithm will produce a valid path in the graph with length $\ell$ in a distributionally correct way — that is, the lengths of the paths drawn will be Poisson distributed (in our case), and paths of any specific length are equiprobable with other paths of that same length.

Both of the estimate for paths passing through at least one of some set of potential observations and our DRAWPATH algorithm are reasonably efficient. Matrix multiplication can be performed in polynomial time (even naïve implementations are $O(n^3)$), so even under simple implementations, the estimation metric is cubic for constant-sized maximum length values, and the DRAWPATH algorithm runs in expected time $O\left(E\{\ell\}^2\cdot|V|^3\right)$, where $E\{\ell\}$ is the expected length of a path and $|V|$ is the number of vertices in the graph. In our implementation, the length factor is reduced from quadratic to linear by eliminating redundant matrix multiplications at the expense of storing all $L^i$ path count matrices created in during the first iteration and using the appropriate matrix during subsequent iterations as $i$ is relaxed. Indeed, we store and expand this list of matrices throughout the estimation process so that the computation for a given $k$ is only ever computed once. The cubic factor in the size of the vertex set can be reduced marginally through the use of efficient matrix multiplication routines [15].

It is worth noting that these two methods can use any length distribution, and they do not rely on the Poisson distribution. Our method merely needs some domain-relevant way of computing $\Pr\{k\}$. Additionally, there are relatively straightforward ways to bias the draw using externally provided edge transition probabilities to replace the $c_b$ vector. These two notions allow for the possible incorporation of a more sophisticated aggregate behavior models, though this is not explored in this paper.

## V. OPTIMIZING POTENTIAL OBSERVATION LOCATIONS

To begin the process, we must determine the best location to place our hypothetical observation points within the graph. We are interested in the best *observational coverage* of the graph — the potential locations for observations in the graph that maximize the probability of observing someone. To be effective, this should take the graph structure into account, as well as the collective behavior model.

From our discussion in the previous section, we know that we can estimate the probability that a path drawn from our collective distribution passes through at least one of some subset of potential observations fairly efficiently using path-count matrix calculations. For this paper, we use a simple evolutionary algorithm (EA) to optimize the locations of a fixed number of possible observation points using our estimate of the probability of observing paths as the objective function.

Our algorithm encodes fixed-length individuals as potential positions for the observations. Out of a population of 20 individuals, one is selected at random to be cloned and mutated. It replaces the worst individual in the population (chosen uniformly at random if there is a tie) if it has at least as high an observational coverage. The number of mutations it undergoes is binomially distributed with the $n=|V|$ and success probability of $\frac{1}{2m}$, where $m$ is the number of potential observations. With equal probability, a mutation event will result in a swap of a position for some other position in the graph *or* a swap of a position for one of its neighboring vertices. The EA is run for 500 steps. These parameters result from parameter tuning, and we make no claim as to their quality except to say that in all cases the EA made progress and appeared to converge to a local optima.

The West Laramie street network graph contains 143 vertices, and during the optimization process, the evaluation heuristic considered paths of length 1 to 15. In spite of this, there are over 7 million paths of such in this graph.

To get a sense for how observational coverage can scale, we ran our EA for several cases of $m$ observations points, $\{128, 64, 32, 16, 8, 4, 2, 1\}$. There were 25 independent random trials conducted for each group. In each trial, we recorded the best observation points discovered by the algorithm. The graph below illustrates these results. In the graph, the points represent the mean best solutions found of the independent runs of the EA, and the whiskers represent the 95% confidence window for each group. We plot a trend

 ISBN: 978-1-61208-251-6 

line through these points to give the reader a sense for how observational coverage scales as the number of observation points is decreased.

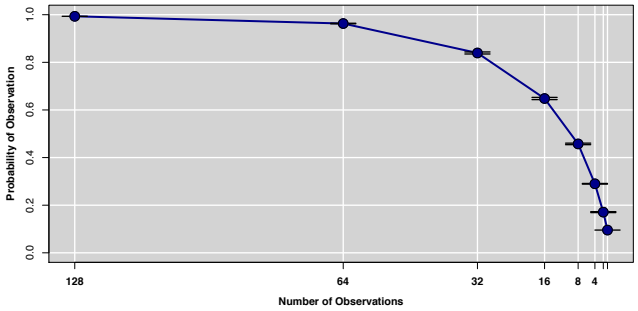



Figure 1. The best observational coverage results for various number of observation points. The points show the mean of 25 independent trial EA runs, and the whiskers show the 95% confidence window for each group.

In each experimental group, the best results were taken and verified as follows: The DRAWPATH algorithm described above was sampled 2000 times and the number of paths that contained at least one of the optimized observation locations was counted. This probability was within the statistical proportions margin of error ($\alpha = 0.05$) of the estimated heuristic based on path-count calculations. Additionally, the path-count estimate for the observational coverage for paths of length 15 or smaller was compared to the same heuristic for path lengths between 1 and 100 with no statistical differences.

Figure 1 illustrates that the observational coverage scales quite well for this graph. Clearly such coverage will differ greatly depending on the graph and the collective behavior model, it is also evident that a great deal of information about individual behaviors in this graph can be available with just a few observation points. Indeed, just two observation points are needed to cover just over 17% of all paths. The location of these observation points is shown the graph below. For the remainder of the paper, we assume that our algorithms have access to observations made at *only* these two points. We refer to these two observations points as $E$ (the east-most point) and $W$ (the west-most point).

## VI. DISTRIBUTION LEARNING

The goal of the first stage of our method is to combine the collective behavior model with hypothetical observations to develop a better understanding of a potential entity movement pattern. While one might simply try to compute the posterior path length distribution when the specific observations are made, we feel it is more constructive to think of this as part of the more general learning process for at least three reasons. First, such computations are likely to be somewhat time consuming estimations. As a result, it is beneficial to develop prior estimates based on hypothetical observations so that sampling can be done relatively

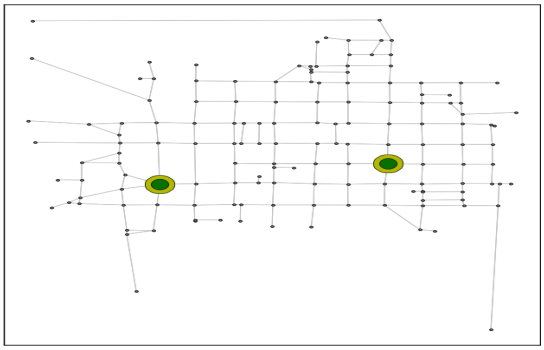


Figure 2. The West Laramie street network with the potential observation points highlighted. These two points represent the best pair of observation locations for coverage discovered by our EA.

quickly in specific cases. Second, prior computation of the posterior distribution facilitates useful analysis of potential observations. These two reasons are justified further in the next paragraph. Finally, when we base our collection of hypothetical observations on historical information, we can consider collections of known observations as a way of encoding aspects of aggregate information about the population. In that sense, it *is* a part of the learning process.

The idea of "hypothetical observations" is rooted in our perception of how aggregate information may be acquired for realistic problems. First, a reasonable scenario is one in which the locations of the observation points are established up front. For example, cameras are specifically located and do not move. If the number of such potential observation points is relatively small, we can pre-compute the posterior path length distribution for many important combinations of observations. Second, we can use a posterior path length estimation algorithm to help with analysis and related, subordinate problems to the main problem. For example, we might examine a variety of potential observation points to see how they affect the length distribution, or use some of the underlying graph theoretic components from the estimation phase to help determine their ultimate placement.

Bayes theorem gives the posterior path length distribution:

$$\Pr\{k|M\} = \frac{\Pr\{M|k\}\Pr\{k\}}{\Pr\{M\}}$$

The value of $\Pr\{k\}$ can be obtained from the historical path lengths as described above. Since $\Pr\{M\}$ is the same value in all cases that use the same observations, one can simply normalize after computing a sufficient number of cases for $k$. One can enumerate cases of $k$ up to some large number, terminating early when it is likely most of the probability distribution has been accounted for. The way we do this is to check when the running ratio described below for each $k$ drops below some small value (e.g., 0.001).

$$\frac{\Pr\{M|k\}\Pr\{k\}}{\sum_{j=1}^{k}\Pr\{M|k\}\Pr\{k\}}$$

 ISBN: 978-1-61208-251-6

We modify the DRAWPATH algorithm discussed in Section IV for drawing fair paths from the collective model to help us compute the condition probability given observations. Here it is assumed that the observations are correct and ordered. That is, if we observe an individual at some vertex $v_a$ then later at $v_b$ but never at $v_c$, then the individual really was at vertex $v_a$ before passing through $v_b$, and he or she never passed through vertex $v_c$. This latter point is important: All observation points contain information about paths since *failing* to see an individual at the point is not the same as no knowing whether the individual visited the point. Consequently, conditioning the path draw based on a negative observation is straight forward: We simply remove the vertex and all in-coming and out-going edges prior to analysis, as we did when we estimated the observational coverage probability.

Modifying the algorithm to consider the ordered positive observations of an individual is more challenging. First note that the algorithm above can be easily altered so that it considers potential end points of the path. To incorporate conditional observations, we combine this idea with a notion we refer to as "*path templates*".

### A. Conditioning the Path Draw End Points

Given the start and end vertices of the path, we can simply replace the third step of the PATHDRAW algorithm above with the specific cell value in the $L^\ell$ matrix indicated by the start and end vertex pair. If only the start vertex is available, we can replace the summation in that step with summing the row indicated by the start vertex. Such a value represents all the unique paths of the given length that start at that vertex. Likewise, if we receive just the end vertex, we can sum the column indicated by that vertex — which is the count of paths of a given length that end in that vertex. The rest of the algorithm remains unchanged.

### B. Path Templates and Piecewise Path Draws

Crucial to understanding our sample estimation methods is the concept of a *path template*. Let us temporarily assume we are given not just observations and a length, but we are also told precisely where in the path the observations where made. We can represent such a path by substituting the unknown vertices with special "wildcard" symbols and treat the entire path as a kind of rule or *template*. For example, given a path length of 6 and observations at vertex $a$ then $b$, one partially informed path is: $\langle ..a.b.\rangle$. A path *matches* such a template if it is of length 6 and goes through vertex $a$ in its third position and vertex $b$ in its fifth position.

Such a template can be broken up into three subpaths: $\langle ..a\rangle$, $\langle a.b\rangle$, and $\langle b.\rangle$. We can then draw uniformly from each subpath using PATHDRAW as discussed above and assemble the complete path after the fact. Such a piece-wise process allows us to draw a path $p$ uniformly from the set of paths that match a specific template. Note that we can also precisely and efficiently compute the number of paths that a template matches by multiplying the number of matches of each of the subpaths.

### C. Putting it Together: Estimating $Pr\{M|k\}$

If path templates were independent of one another with respect to the paths they match, one needn't even use DRAWPATH: Simply use the method just described to compute the total number of paths that match a given template, and instead sample templates of a given length. Knowing this, in addition to the total number of paths of a given length (unconditioned by observations) gives an easy and efficient way to compute precisely the probability that of a template match given path length. This reduces the problem to sampling amongst templates rather than paths.

In principle, templates can be produced in a number of ways. One method is to enumerate them given observations. While this scales combinatorially with the length of the path and number of observations, it is still quite manageable for relatively small paths with sparse observations. Alternatively, we can produce a template uniformly at random by using a simple and efficient shuffling method.

Unfortunately, two different templates may match the same path so the templates are not independent of one another. We can use the *inclusion-exclusion* principle [16] to adjust these match counts if we have a means of computing the union of the match sets represented by two templates. This is straightforwardly accomplished by *merging* templates: If two fixed positions do not match then the union of the templates produces 0 matches, otherwise the merged template is produced by retaining all fixed positions between the two templates. For examples $\langle a...b.c\rangle \cup \langle a.d...c\rangle = \langle a.d.b.c\rangle$. However, overall this computation can be quite costly in the worst case since there can be a combinatorial number of levels of such unions. It may be that only a few levels are needed to get a reasonable approximation, and there may be ways of using piece-wise DRAWPATH to reduce the overhead of full enumeration. We would like to explore this possibility in the future.

We store the templates that with non-zero numbers of paths that were discovered during this process, along with the number of paths they match (*match count*) as explicitly computed from the path-count matrices as described above. This is used during resampling, as we shall see.

Consider the West Laramie street network with two fixed observation points as determined by our optimization discussed in Section V, $E$ and $W$. Since positive observations are ordered but negative observations are not, there are five cases for learning. Observation(s) at:

i) $E$ then $W$;
ii) $W$ then $E$;
iii) $E$ but *not* $W$;
iv) $W$ but *not* $E$;
v) neither point.

Since there are only a handful of such cases, we are free to learn these distributions *a priori*. In more complex scenarios,

 ISBN: 978-1-61208-251-6

it may be necessary to learn only the a few observational scenarios, then add new cases as they occur.

## VII. Path Reasoning Given Sparse Observations

Once we have learned a catalogue of distributions based on potential observations, we can use this information to resample paths given specific, actual observations. In much the same way as the distribution is sampled in the first place, this problem includes drawing a length, then a template, then a path from that template. Because of the dependence problem between templates above, we do not use the sample to estimate probability, but instead assign the probability of a path directly from the posterior distribution. This process is described in more detail below.

**Algorithm 2** PathResample (observation set $M$, graph $G$, posterior path length distributions, templates & match counts, maxSamples)

```
Initialize path list to an empty list
for i = 1 to maxSamples do
    Lookup posterior {k|M} distribution
    Draw path length according to Pr{k|M}
    Lookup recorded templates & match counts
    Randomly select template proportionate to match count
    for each unspecified subpath in the template do
        Randomly select valid subpath using piecewise draw
    end for
    if path is unique then
        Assign the path probability Pr{k|M}
        Store path and probability in path list
    end if
end for
Remove from path list all but the 5% most probable paths
return path list with associated probabilities
```

The list of paths recovered from the resampling procedure is used in two ways. In each of the five observational cases, the graph can be plotted with the resampled paths overlaid for visualization. Second, a quantitative measure of *informativeness* of the resampled path can be estimated based on an edge count frequency histogram: All edges are treated as separate bins, and an edge is counted once for each of the resampled paths in which it participates, then the entire list of edges is normalized to create a distribution of edge usage in the resampled paths. Given a baseline condition that equiprobable edges implies no information, we compute the distance of this edge count frequency histogram to a vector of length $|V|$ with all values equal to $1/|V|$, where $|V|$ is the number of vertices in the graph. These visually and intuitively confirm the idea that the observations can provide significant context about an entity's behavior

The results presented in Table I indicate that having any positive observations at all provides a substantive amount of information over and above no detection. Figures 3, 4, and 5 we present three graphs that are visually indicative of these results, cases $i$, $iii$, and $v$, respectively.

Table I
Distance measures between uniform distribution of edges and the edge count frequency histograms constructed from resampled paths for each learned distribution.

| Case # | Case Description | Distance |
|---|---|---|
| $i$ | Observation at $E$ then $W$ | 0.2206 |
| $ii$ | Observation at $W$ then $E$ | 0.2289 |
| $iii$ | Observation at $E$ but *not* $W$ | 0.2403 |
| $iv$ | Observation at $W$ but *not* $E$ | 0.2223 |
| $v$ | Observations at neither point | 0.0949 |

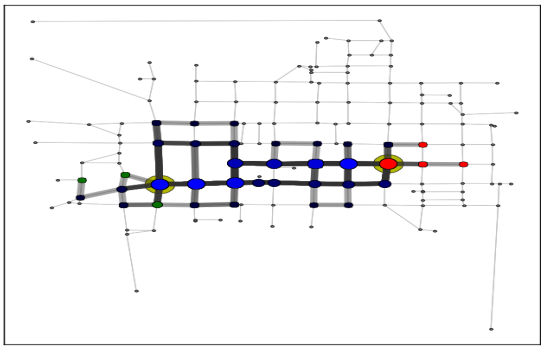


Figure 3. The 5% most probable resampled paths for an individual on the West Laramie street network, under the condition that he/she was observed first at the east-most (E) observation point then the west-most (W).

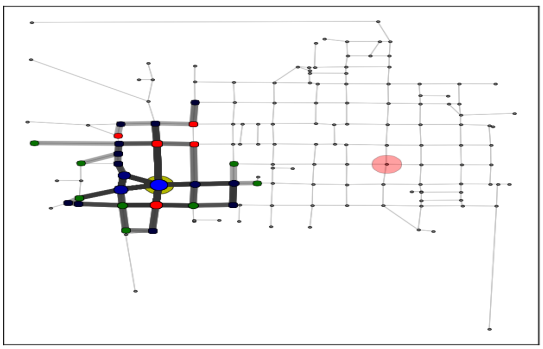


Figure 4. The 5% most probable resampled paths for an individual on the West Laramie street network, under the condition that he/she was observed at the east-most (E) observation point but not at the west-most (W).

## VIII. Conclusions

In this work, we adopt the position that, because the structure of a graph itself contains a great deal of information, one can use relatively simple graph theory and probability techniques to learn a lot about potential behavioral patterns of entities operating within the graph even under quite

 ISBN: 978-1-61208-251-6

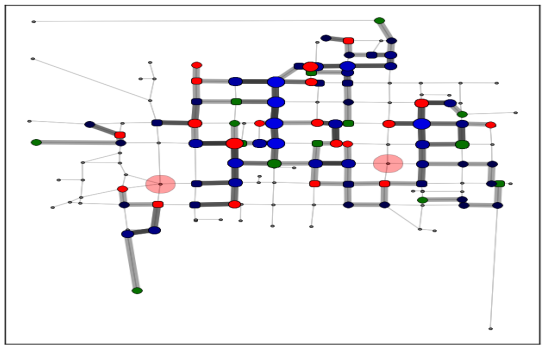


Figure 5. The 5% most probable resampled paths for an individual on the West Laramie street network, under the condition that he/she was not observed at either location.

spartan conditions: There is little *a priori* information available about collective movement and there are few specific observations available.

We presented a simple technique for estimating the observational coverage of a set of fixed observations — the probability that a path drawn from our prior path distribution will pass through at least one of the observation points. This method was used by a basic evolutionary algorithm to find near optimal positioning of fixed potential observation points within the graph. As it turns out, for the West Laramie street network graph, the observational coverage scales nearly logarithmically as the number of observational points is reduced. Indeed, just two fixed positions can cover nearly a fifth of all paths in the graph.

We also presented an efficient way to draw a path from that prior distribution in a probabilistically correct way. We expanded this algorithm to permit the path draw to incorporate conditions associated with ordered observations. We used this algorithm to construct five models of to match potential entity movement in a graph using five cases of hypothetical observations. These models were used to resample paths from the posterior path distributions to produce the most probable paths an entity might have taken given the different observational scenarios. Both quantitative and qualitative evidence was provided that such observations provide a great deal of contextual information about possible entity behavior.

In the future, we are interested in expanding our collective model of behavior to incorporate more sophisticated prior bias information. For example, there may be ways to incorporate time by combining the use of the Poisson length distribution with rate-of-travel and edge cost information over fixed time constraints. Additionally, we can incorporate edge transition probability information into the DRAWPATH algorithm to bias the draw based on a learned aggregate model. This may be facilitated but our complementary and parallel efforts to make use of biased random walker models as a means constructing edge transition probabilities.

## REFERENCES


[1] A. Asahara, K. Maruyama, A. Sato, and K. Seto, "Pedestrian-movement prediction based on mixed markov-chain model," in *Proceedings of the* $19^{th}$ *International Conference on Advances in Geographic Information Systems*. New York, NY, USA: ACM Press, 2011, pp. 25–33.

[2] G. Vaca-Castano, A. Roshan Zamir, and M. Shah, "City scale geo-spatial trajectory estimation of a moving camera," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2012.

[3] K. A. De Jong, *Evolutionary Computation: A Unified Approach*. MIT Press, 2006.

[4] T. Hägerstrand, "What about people in regional science?" *Papers in Regional Science*, vol. 24, no. 1, pp. 6–21, 1970.

[5] H. Miller, "Necessary space- time conditions for human interaction," *Environment and Planning B: Planning and Design*, vol. 32, no. 3, pp. 381–401, 2005.

[6] A. Crooks, C. Castle, and M. Batty, "Key challenges in agent-based modelling for geo-spatial simulation," *Computers, Env. and Urban Sys.*, vol. 32, no. 6, pp. 417 – 430, 2008.

[7] A. Willis, N. Gjersoe, C. Havard, J. Kerridge, and R. Kukla, "Human movement behaviour in urban spaces: Implications for the design and modelling of effective pedestrian environments." *Environment and Planning B Planning and Design*, vol. 31, no. 6, pp. 805–828, 2004.

[8] D. Helbing, P. Molnár, I. Farkas, and K. Bolay, "Self-organizing pedestrian movement," *Environment and Planning B*, vol. 28, no. 3, pp. 361–384, 2001.

[9] B. Hillier and S. Iida, "Network and psychological effects in urban movement," *Spatial Info. Theory*, pp. 475–490, 2005.

[10] W. Spears and S. Prager, "Evolutionary search for understanding movement dynamics on mixed networks," *GeoInformatica*, pp. 1–33, 2012.

[11] A. O'Connor, A. Zerger, and B. Itami, "Geo-temporal tracking and analysis of tourist movement," *Mathematics and Computers in Simulation*, vol. 69, no. 1-2, pp. 135–150, 2005.

[12] R. Church and C. VELLE, "The maximal covering location problem," *Papers in regional science*, vol. 32, no. 1, pp. 101–118, 1974.

[13] A. T. Murray, K. Kim, J. W. Davis, R. Machiraju, and R. Parent, "Coverage optimization to support security monitoring," *Computers, Environment and Urban Systems*, vol. 31, no. 2, pp. 133 – 147, 2007.

[14] N. Xiao, "A multiobjective evolutionary algorithm for surveillance sensor placement," *Environment and Planning B: Planning and Design*, vol. 35, pp. 935–948, 2008.

[15] T. H. Cormen, C. E. Leisserson, R. R. L., and C. Stein, *Introduction to Algorithms*, 2nd ed. MIT Press, 2001.

[16] C. M. Grinstead and L. J. Snell, *Introduction to Probability*. American Mathematical Society, 2006, version dated 4 July, 2006.


 ISBN: 978-1-61208-251-6 

# Advanced Scientific Computing and Multi-Disciplinary Documentation for Geosciences and Archaeology Information

Claus-Peter Rückemann
*Westfälische Wilhelms-Universität Münster (WWU),*
*Leibniz Universität Hannover,*
*North-German Supercomputing Alliance (HLRN), Germany*
*Email:* ruckema@uni-muenster.de

***Abstract*—This paper presents the results from the implementation of Integrated Information and Computing System components based on a new universal framework using multi-disciplinary documentation and advanced scientific computing for enabling long-term use of information in geosciences and archaeology. The solution enables the creation of general long-term knowledge resources, which will stay highly efficient, most flexible, extensible, and economic for any kind of use. The core is an architecture consisting of a knowledge resources base, application resources, and sources. The knowledge resources can contain and refer to any object. For providing a universal integration of objects and in order to improve the quality of the data, a long-term integration of structured objects and universal classification has been developed. This way, any content and context can be described and used for advanced application scenarios, deploying high end compute and storage resources for any processing and computing tasks. Due to a flexible collaboration framework the implementation allows dynamical use with information systems and supercomputing resources for any kind of workflow and application scenario, integrating knowledge from natural sciences and humanities.**

***Keywords–Integrated Systems; Information Systems; Documentation; Advanced Scientific Computing; Classification; Archaeology; Geosciences; High Performance Computing.***

## I. INTRODUCTION

With common collections of information and data today, even if widely accessible, we are missing reliability, validation, and long-term sustainability. This is resulting from principle problems of implementation: There are no foundations of a suitable long-term strategy, documentation, tools, and resources. This demands static and dynamical components in all parts of an implementation. Content has to be developed for long periods of time. This includes new research, including historical information, and extending multi-disciplinary references. The focus question for documentation, operating on information and computing is: How can complex systems be built, developed, and extended over the necessarily long periods of time? With the application components, e.g., information system components like databases, mostly form monolithic and even proprietary blocks. Their life cycle is mostly much shorter than long-term content development. On the side of computing challenges it is possible to create solutions for very perishable present resources. Any more complex problem cannot be considered "solved" for future architectures and applications. Many information and knowledge resources cannot be used without the original context, e.g., computing resources any more. Up to now context of information science cannot be described by common means to a reasonable extend. This leads to another essential question: What information and knowledge on content and context can be preserved for medium- and long-term usage when the complexity of an overall system will be unpredictably high? The long-term strategy created here is based on an implementation architecture, which includes long-term knowledge resources with the resources and development. In this paper we concentrate on the archaeological and geosciences topics being part of the knowledge resources. The foundations should enable the essential processing of archaeological, geoscientific, geophysical, geological, spatial and other data as well as a thorough documentation of all aspects of content and context and the exploitation of advanced scientific computing methods and resources for maximum flexibility. This paper will describe the potential of Integrated Information and Computing Systems (IICS) being based on multi-disciplinary documentation for this purpose. Anyhow, it can only describe a tiny fraction of the multitude of possible features.

This paper is organised as follows. Section two describes the motivation and Section three lists aspects of related work. Section four introduces the architecture and implementation for the IICS created. Section five describes the aspects of the long-term strategy. Section six discusses an implementation case study and basic mechanisms for a geoscience-archaeology IICS. Section seven presents an evaluation including processing, computing, and classification aspects. Section eight summarises the conclusions and future work.

## II. MOTIVATION

There are no frameworks providing the necessary concepts and features for integration of data, workflows, knowledge and computing resources, and operation. For example, a lot of advanced geoscientific processing cannot be reproduced after a few years, even if the data and results are still

 ISBN: 978-1-61208-251-6 

existing. Therefore means for extended long-term interpretation and analysis are missing. It is a huge challenge that, besides data creation not being able to support sufficiently comprehensive documentation, the widely used technology, e.g., document formats, Uniform Resource Locators (URL), and Web Services are not persistent over longer periods of time, e.g., for static objects file formats do change, for applications the implementations will change, and for services the features will be modified. Therefore information structures built from such technologies will become inaccessible. Long-term knowledge creation cannot rely on this. From the complex systems' point of view any of those building elements are not suitable for describing objects and creating long-term knowledge resources. Anyhow the original sources and building elements are needed for documentation of the original content and context. Therefore knowledge creation has to separate the essentials of knowledge from technology, resources, and other tools while at the same time respecting their importance. Even worse , that workflows, algorithms, resources and their management cannot be guaranteed for long-term availability. The topic is very complex and experiences with long-term knowledge creation are out of scale of the time interval of most researchers. Especially, it has been found to be less difficult for groups with a strong background of classical academic education to understand the problem itself, than it is for groups with a "technical-only" background to realise the multiple benefits of classification. A basic example of long-term creation of knowledge and the implementation of applications building on these resources are presented in the following sections. It shows a scenario that can hardly be managed with other available methods and concepts in a comparable way.

## III. Related Work

There is no wider concept and implementation known comparable to the solution presented, described and implemented here. Nevertheless, there are concepts for components, implementations, and terminology. Previous work [1], [2], [3], [4], [5] has delivered important basic concepts and components, e.g., Integrated Information and Computing Systems, dynamical components, and taxonomy. Taxonomy is the science and practice of classification. An important facetted classification is the Universal Decimal Classification (UDC) [6]. According to Wikipedia currently about 150000 institutions, mostly libraries, are using basic UDC classification worldwide [7], e.g., with documentation of their resources, library content, bibliographic purposes, for digital and realia objects. This is mostly restricted to publications and references but not of general knowledge and applications. Some aspects can be studied from the goals of knowledge discovery [8], which is becoming increasingly important. Other aspects are handled with search algorithms, which currently are still very primitive regarding knowledge creation and usage.

## IV. Architecture and Implementation

As far, it is not commonly possible to treasure content currently used for being preserved in order to create really long-term usable content. Even much more difficult that an implementable solution for any form of long-term context is even in wide distance. In general, only a very small percentage of disciplines and researchers are familiar with knowledge classification and applications. The more, multidisciplinary classification is currently only in the focus of third parties. Operational resources and features are considered to be short-term issues whereas information, knowledge, and respective resources and features must be considered of long-term significance. Case studies showed that long-term development requires a strong sustainability of content, context, and computation. This means, the most important part of these systems is the knowledge resources containing the content and context documentation to any extent necessary for describing the activities and isolating the perishable components for context documentation. A solid classification cannot be done automatically. The more, it cannot be done automatically for use with IICS. Anyhow, in fact that different views are possible, it is reasonable to have classification view from the origin, from main disciplines or from the developers in order to increase the quality of references. The architecture respects these conditions. The following sections explain how a successful implementation of an integrated system can be created and operated using knowledge resources and classification for information system usage.

### A. Architecture for documentation and development

The architecture implemented for an economical long-term strategy is based on different development blocks. Figure 1 shows the three main columns: Applications resources, knowledge resources, and originary resources.

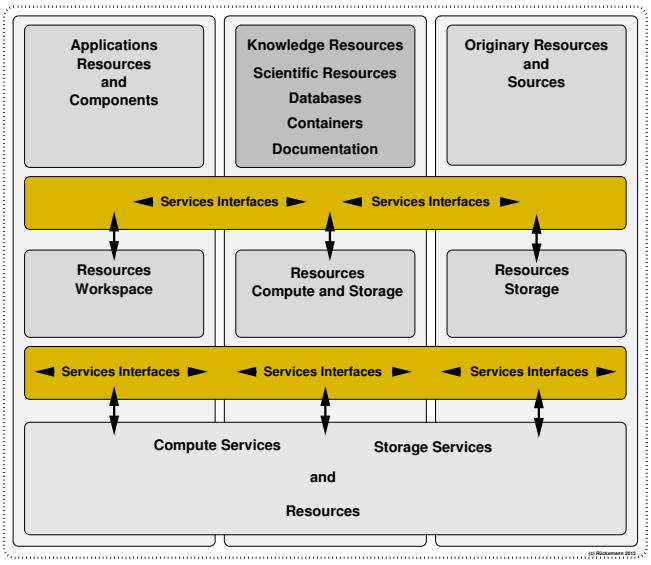



Figure 1. Architecture: Columns of practical dimensions.

The central block in the "Collaboration house" framework architecture [5], are the knowledge resources, scientific resources, databases, containers, and documentation (e.g., LX [1], databases, containers, list resources). These can be based on and refer to the originary resources and sources (photos, scientific data, literature). Application resources and components (Active Source, Active Map, local applications) are implementations for analysing, utilising, and processing data and making the information and knowledge accessible. These three blocks are supported by services interfaces. The interfaces interact with the physical resources, in the local workspace, in the compute and storage resources the knowledge resources are situated, and in the storage resources for the originary resources. All of these do allow for advanced scientific computing and data processing as well as the access of compute and storage resources via services interfaces. The resources' needs depend on the application scenarios to be implemented for user groups.

### B. Components: Applications, knowledge, and sources

The main information, data, geo-referencing, and algorithms for all presented components and examples are provided by the LX Foundation Scientific Resources [1]. This deploys the structure and classification of objects necessary for a reasonable implementation. Besides the LX structure the already established Universal Decimal Classification (UDC) [6] has been integrated for objects [2] as it provides a hierarchical and multi-lingual, faceted classification for any topic and allows implementing a faceted analysis with enumerative scheme features, as well as to create new classes by using relations and grouping. In multi-disciplinary object context, this empowers to use workflows combining keywords, enumerative concepts and full-text analysis with a faceted analysis. Besides the academic, industrial, and business application scenarios in focus of the GEXI collaborations' case studies [3] it is an important factor to integrate the necessary documentation and computing facilities with systems like an Universal IICS (UIICS). An implementation of interfaces for using structure and classification with appropriate Archaeological IICS system components has been created for several simple (SAMPLE, COLLECTION, CONTEXT, DISCIPLINE) and slightly more complex workflows (CONNECT, REFERTO-TOPIC, REFERTO-SPATIAL, VIEW-TO, VIEW-FROM) [5].

For the topics and content discussed here, geoscientific and archaeological information and processing are the core content. Data in this context necessarily includes applications and algorithms. Besides the above implemented features, it is optional to support any visualisation tool, processing algorithms, cartographic and mapping features and many more tools from secondary sciences, e.g., spatial algorithms and components, UDC (1-0/-9). These features can be used with the objects in any way that will be necessary to describe data and automate workflows.

### C. Classification, keywords, and interfaces

The interfaces allow to use the various resources. A central element is the classification and structure of the knowledge resources as it increases the flexibility of the long-term development. Table I compares some features of classification and keywords used for object description.

Table I
UDC CLASSIFICATION AND KEYWORDS COMPARISON.

| *UDC* | *Keywords* |
|---|---|
| Internationalisation | Methodical support, partial internationalisation |
| Codes | Code table support |
| High level of detail | Medium level of detail |

Interfaces can be used in order to access and use objects. This includes filtering, combination, workflow and data processing and so on. In summation, this allows the integration of all data, objects, and resources available: scientific and discipline data, lexicographical and bibliographical data GPS data, geospatial information, processing algorithms, executable software, and many more, including realia objects. It means, we need to integrate multi-disciplinary information, allowing different views on the same context and allow even different paths for exploring knowledge.

## V. LONG-TERM STRATEGY

Whereas looking from inside a traditional discipline, information seems to be complete and appears to increase slowly. On the other hand there is huge complementary information that cannot be described isolated by one discipline and tendency is increasing over the time and complexity. Table II presents the result of a reasonable categorisation that has been found from practicing the knowledge resources creation and use for several decades. It shows a more detailed compilation of categorised features and components for an expected actuality time range. In this context, the goal for long-term means $> 50$ years, medium-term $> 15$ years, and short-term $< 15$ years.

Table II
TIME-RANGE GOALS WITH IICS COMPONENTS (SELECTION).

| *Long-term* | *Medium-term* | *Short-term* |
|---|---|---|
| Knowledge | Applications | Context |
| Containers | Interfaces | Sources |
| LX Resources | DOI, URN | URL |
| UDC | Converters | Media |
| Keywords | Active Source | Converters |
| Virtualisation information | Storage resources | Computing resources |
| Algorithms | Distributed services | Compiler, Executables |
| Content | Virtualisation | MPI, OpenMP |
| Context information | Complex implement. | Batch systems |
| Relations & references | Application features | Web Services |
| Internationalisation | OS features | Communication |
| Processing & workflows | Library features | Middleware |

These components, described by a representative selection in Table II, can cover all aspects of knowledge creation, application, and system implementation. For example, with

 ISBN: 978-1-61208-251-6

the implementation, the resources and containers are consisting of thousands of pages. For the presentation within the following sections a small excerpt of the objects and classification can be shown. The long-term objects must be able to contain the essential knowledge, even as medium- and short-term objects cannot be preserved or made persistent as, e.g., DOI (Digital Object Identifier), URN (Uniform Resource Name), and URL (Uniform Resource Locator) will vanish and context and sources made fade away as well as OS (Operating System) features used. Therefore we have to distinct between the real instance of a DOI and URL or a context situation and a descriptive reference of these objects. These descriptive references can contain as much information and knowledge as possible (for example DOI, URL, context description, sources).

## VI. Mechanisms Case Study

A filter chain can be used to compute resulting object sets. Based on the available system the following steps can be separated, in an example from geosciences and archaeology:

- Select topic from knowledge base (volcanology).
- Select region from results (Europe, Caribbean).
- Select volcano from results (Vesuvius, La Soufrière).
- Select object entries (geosciences and archaeology).
- Select media objects and references.
- Select application resources and interfaces.

### A. Workflows and algorithms

The knowledge resources block is the central resource in the long-term strategy. The knowledge resources can contain any kind of content. Application components can be migrated into the knowledge resources for documentation purposes and re-use. The services can access archived and historical data as well as live data and feed it into the workflows. Services interfaces allow to build complex workflows using arbitrary algorithms. The knowledge resources can be accessed from applications, which will extract suitable information and trigger the use of compute and storage resources. Objects can be selected by any algorithm, e.g., combinatory, search, and filter algorithms. The results can be delivered to a defined location or service.

### B. Resulting cross-links calculated

The knowledge resources organisation allows to create all structures and references possible for information science purposes, besides content and taxonomy. A small subset of knowledge space objects is less suitable for providing the necessary depth for building facets and referring to appropriate multi-disciplinary objects. Therefore a basic example needs an extended knowledge environment. Table III shows an excerpt of the references and cross-links calculated from the knowledge resources for a "Vesuvius" object with some applications and originary resources, media information and illustrative examples. The secondary references (symbol: •-○) for an object are calculated from the respective reference matrix in the knowledge resources.

Table III
Example cross-links with secondary references for "Vesuvius" object and generated views (excerpt).

| *Knowledge Resources* | *Applications Resources* | *Media Inform.* | *Originary Resources* |
|---|---|---|---|
| **Vesuvius** (LX Resources) | actmap | photo | samples |
| •-○**volcano** | | object | [on-site] |
| •-•-○ **climatology** | | object | – |
| •-•-○ **magma chamber** (see Listing 5) | | object | [on-site] |
| •-•-•-○ **Thrihnukagigur** | | links | [on-site] |
| •-•-•-•-○ **Þríhnúkagígur** | Transcription | links | [on-site] |
| •-○**super volcano** | [interfaces] | object | [on-site] |
| •-•-○ **Yellowstone** (see Figure 3) | | object | [on-site] |
| •-•-•-○ **Liparite** | | object | samples |
| •-•-•-○ **Rhyolithe** | | object | samples |
| •-○**Soufrière** | | object | samples |
| •-○**La Soufrière, Guadeloupe** | Movieview | video | samples |
| •-•-○ **Lava sand, black** | | object | samples |
| •-•-•-○ **Plage de Grande-Anse** | | object | samples |
| •-○**Mt. Scenery, Saba** | Topicview | object | samples |
| •-○**archaeology** | Photoview | photo | – |
| •-•-○ **Pompeji** (see Figure 2) | | object | museum |
| •-•-○ **Herculaneum** | | object | museum |
| •-•-○ **Hephaistos** | | object | – |
| •-•-○ **Vulcanus** (see Figure 4) | | object | – |
| •-•-•-○ **Puteoli** | | object | – |
| •-•-○ **Zipacná** | | object | – |
| •-•-○ **Cabrácan** | | object | – |
| . . . [cut] . . . | Objectmap | | – |
| •-○**Volcano compendium** | Volcanomap | | |
| •-•-○ **VNUM** | | | |
| •-•-○ **Location** (see Figure 5) | | | |
| •-○**Archaeological compendium** | | | |
| •-○**Geological table** | | | |
| •-○**Stone table** | Stonetab | object | samples |
| •-○**Mineral table** | Mintab | object | samples |
| •-•-○ **Periodic table of elements** | Elemview | object | samples |
| | Mapview | image | – |
| . . . | Satelliteview | service | – |
| | | | – |
| | Index | service | – |
| | Dynasearch | service | – |
| | Dynacompute | service | – |
| | Dynaprocessing | service | – |
| | Dynaindex | service | – |
| | Dynatypeset | service | – |
| | Dynastat | service | – |
| [compute and storage resources] | [workstation] | | |

(Aerial image data, e.g., source: Google Maps)

The cross-links are extracted from the long-term knowledge resources and are usable with available data by any application components. Available example instances from the application components set, media resources, and originary sources (realia) are listed. In this context cross-linked subentries, keys and, for example, in case of volcanic features the classification (UDC) and the Volcano Number (VNUM) can be integrated in the analysis for computation and processing via the reference to the appropriate container, the volcano compendium. Applications and interfaces (e.g., Active Source, actmap [4]) can access any available data, as creating index compilations, retrieving online maps and

 ISBN: 978-1-61208-251-6 

satellite imagery. Media are, e.g., video data and photo images. Originary sources of objects are, e.g., realia, in the case with the volcanic material here, the objects are volcanic samples, respective stones. These objects are referred to in an appropriate archive, e.g., a collection, library or museum. The bottom row shows how the services and resources are used (e.g., local workstation access, special compute and storage resources).

### C. Resulting object classifications and processing

Listing 1 shows a simple UDC-Context sample subentry from the LX Resources. This example expresses a basic "Europe : America" relation.

```
UDC-Context:(4):(7)
```

Listing 1. UDC-Context sample (LX Resources).

Classification patterns are suitable to select objects with any algorithm and pattern matching rules. Listing 2 shows a basic grep-filter within a classification set of UDC samples.

```
... | egrep "\([0-9]\):\([0-9]\)" | grep "(4):(7)"
```

Listing 2. Search grep sample within classification (LX Resources).

The result is a set of objects with a context relation between the continents Europe and America. Listing 3 shows an excerpt of an UDC table used with the examples.

```
%%IML: UDC55     :: Earth Sciences. Geological sciences
%%IML: UDC56     :: Palaeontology
%%IML: UDC57     :: Biological sciences in general
%%IML: UDC911.2  :: Physical geography
%%IML: UDC902    :: Archaeology
%%IML: UDC903    :: Prehistory. Prehistoric remains, artefacts, antiquities
%%IML: UDC904    :: Cultural remains of historical times
%%IML: UDC930.85 :: History of civilization. Cultural history
%%IML: UDC"63"   :: Archaeological, prehistoric, protohistoric periods, ages
%%IML: UDC(4)    :: Europe
%%IML: UDC(7)    :: North and Central America
%%IML: UDC(23)   :: Above sea level. Surface relief. Above ground generally.
%%IML: UDC(24)   :: Below sea level. Underground. Subterranean
%%IML: UDC=12    :: Italic languages
%%IML: UDC=14    :: Greek (Hellenic)
%%IML: UDC=84/=88:: Central and South American indigenous languages
```

Listing 3. UDC classification table, English (LX Resources, excerpt).

For a Topicview, corresponding object classes are processed and media objects are computed. Listing 4 shows the core data generated from the filtered media data, processing images for a sample Topicview as listed in Table III.

```
create_archaeology_planet_view_topic.sh \
   volcano_guadeloupe_soufriere_viewto.jpg \
   volcano_guadeloupe_soufriere_viewfrom.jpg \
   volcano_saba_mtscenery_viewto.jpg \
   volcano_saba_mtscenery_viewfrom.jpg
```

Listing 4. Generated core data for Topicview processing.

As the knowledge resources' objects carry references to any kind of detailed processable data, distribution maps and satellite views can be computed and passed on.

### D. Resulting object entries on geosciences and archaeology

The following object entries are excerpts from the calculated cross-links table (Table III). The excerpts contain some, structure, UDC classification, keywords, references, and satellite image reference. The references for the geopositioning are created via classification. They can be used for any purpose where geopositioning will seem interesting. Listing 5 shows an excerpt of an LX Resources object entry [1], "Vesuvius" volcano.

```
Vesuvius [Vulcanology, Geology, Archaeology]:
         (lat.) Mons Vesuvius.
         (ital.) Vesuvio.
         (deutsch.) Vesuv.
         Volcano, Gulf of Naples, Italy.
         Complex volcano (compound volcano).
         Stratovolcano, large cone (Gran Cono).
         Volcano Type: Somma volcano,
         VNUM: 0101-02=,
         Summit Elevation: 1281\UD{m}.
         The volcanic activity in the region is observed by the Oservatorio
         Vesuviano. The Vesuvius area has been declared a national park on
         \isodate{1995}{06}{05}. The most known antique settlements at the
         Vesuvius are Pompeji and Herculaneum.
         Syn.: Vesaevus, Vesevus, Vesbius, Vesvius
         s. volcano, super volcano, compound volcano
         s. also Pompeji, Herculaneum, seismology
         compare La Soufrière, Mt. Scenery, Soufriere
         %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(4+23+24)=12=14
         %%IML: GoogleMapsLocation: http://maps.google.de/maps?hl=de&gl=de&
          vpsrc=0&ie=UTF8&ll=40.821961,14.428868&spn=0.018804,0.028238&t=h&z
          =15
```

Listing 5. Knowledge resources – object entry "Vesuvius" volcano.

The example contains a reference and VNUM for the Vesuvius volcano, various secondary objects, UDC classification, satellite image reference (Satelliteview in Table III). It refers to "Soufriere", "La Soufrière", and "Mt. Scenery". Listing 6 shows an excerpt of the "Soufriere" object entry.

```
Soufriere    [Vulcanology, Geology]:
             A common name for a volcanic feature resulting from the
             french term for \periref{tgt:PeriSulfur}{Sulfur}.
             The name soufriere is used for a volcanic crater
             or other area in combination with solfataric activity.
             The name is mostly used in French speaking regions,
             especially in the West Indies.
             Very well known are, for example:
             La Soufrière volcano, Guadeloupe, F.W.I.
             Soufriere Hills, F.W.I.
             Soufriere St. Vincent, F.W.I.
             Syn.: Soufrière
             s. also La Soufrière, F.W.I., volcano, seismology
             %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(7+23)=84/=88
```

Listing 6. Knowledge resources – secondary object entry "Soufriere".

This secondary object entry, "Soufriere", also refers to the La Soufrière volcano (Listing 7), which itself refers to various data and objects, e.g., satellite image references.

```
La Soufrière [Vulcanology, Geology]:
             La Soufrière volcano, Guadeloupe, F.W.I.
             Volcano Type: Stratovolcano,
             Country: France,
             Subregion Name: West Indies, Caribbean,
             VNUM: 1600-06=,
             Summit Elevation: 1467\UD{m}.
             Syn.: Soufriere
             s. volcano
             s. also Soufriere, F.W.I., lava, lava sand, OVSG
             %%IML: UDC:[911.2+55]:[57+930.85]:[902]"63"(7+23+24)=84/=88
             %%IML: GoogleMapsLocation: http://maps.google.com/?ie=UTF8&ll
              =16.043153,-61.663374&spn=0.003088,0.003262&t=k&z=18&vpsrc=6&
              lci=weather
```

Listing 7. Knowledge resources – secondary object "La Soufrière".

The secondary object entry "Mt. Scenery" (Listing 8) also contains classifications and media, and further data references for the Mt. Scenery volcano on Saba. Extracted examples are volcano type, VNUM, region, status, elevation, and UDC classification views as well as the geo-references, which with this request are used to automatically compute views and distribution maps for classified objects in the result matrix. The classification groups themselves show references to associated objects. The data and media object in a processed reference chain can be used for further

 ISBN: 978-1-61208-251-6

analysis, creating special features. That way, using UDC classifications, e.g., places from a region or context that can be associated with volcanology and associated with archaeological sites can be selected and media objects can be processed and realia referred.

```
Mt. Scenery  [Vulcanology, Geology]:
             Volcano, Saba, Netherlands Antilles, D.W.I.
             Volcano Type: Stratovolcano,
             Country: Netherlands,
             Subregion Name: West Indies, Caribbean,
             VNUM: 1600-01=,
             Volcano Status: Historical,
             Last Known Eruption: in or before \isodate{1640}{}{},
             Summary Elevation: 887\UD{m}.
             %%IML: UDC:[55+56+911.2]:[902+903+904]:[57+930.85]"63"(7+23+24)
              =84/=88
             %%IML: GoogleMapsLocation: http://maps.google.com/maps?f=q&
              source=s_q&hl=en&geocode=&q=mt+scenery,+saba+netherlands+
              antilles,+google+maps&aq=&sll=17.633225,-63.236961&sspn
              =0.048997,0.052185&vpsrc=0&t=h&ie=UTF8&hq=mt+scenery,+saba+
              netherlands+antilles,&hnear=&z=14&lci=weather
             s. also Saba, D.W.I., volcano, seismology
```

Listing 8. Knowledge resources – object entry “Mt. Scenery”.

Dynamical components can even benefit from precalculation and precomputation of objects. This includes precalculated classification and weights (“PreUDC”). The following section presents examples calculated from the above classified objects and the figures are showing the results of selected attributes, including the classification and geo-references used basic visualisation.

### *E. Resulting features selection for cross-links processing*

The following (Figure 2) is an excerpt of the secondary objects computed above for the Caribbean region volcanoes and a selection with UDC “(23)”, “(24)”.

Figure 2. Topicview – volcanoes, La Soufrière (left), Mt. Scenery (right), VIEW-TO (green), VIEW-FROM (blue).

Figure 3 illustrates the computed objects (Topicview), e.g., here volcanic samples after processing, all showing the variety of material from the top of the La Soufrière volcano.

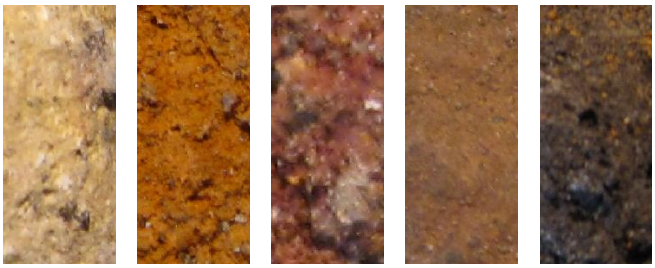


Figure 3. Topicview – related volcanic samples (La Soufrière, 2011).

Any of these objects being part of the resulting matrix for a request, e.g., photos for object entries as well as media data for physically available samples, have been found via references and UDC from the knowledge base. The realia references for the objects refer to a collection where the samples are stored. Further analysis for the samples is available via the knowledge resources. Figure 4 shows the geolocations [1] for computed geoscientific and archaeological object samples on a configurable object map.

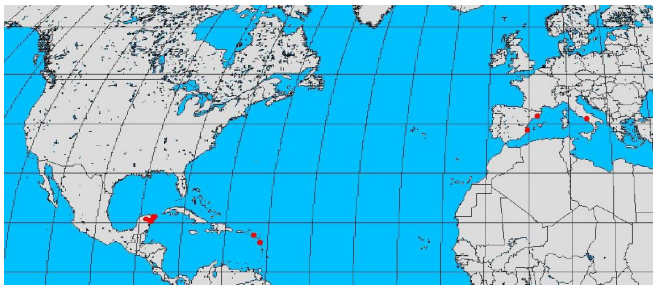


Figure 4. Objectmap – computed map for related objects (red, excerpt).

A sample distribution of volcanic features is depicted in Figure 5. It shows a comparison of volcanic data in a projection identical to the computed Objectmap (Figure 4). Although the knowledge matrix of this example is most complex (Table III), the workflow for producing a view can be specified very easy like for spatial presentation.

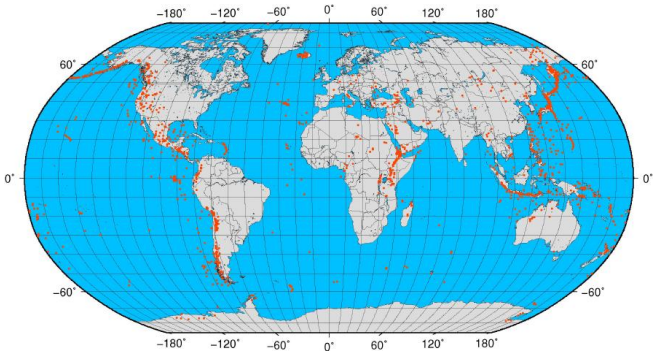



Figure 5. Volcanomap – worldmap of referenced volcanoes (orange).

The map generated with the workflow as described with the case study presents the related objects from the context available in the geophysical research database. It can visualise various aspects of the classified objects. In this case of volcanoes and geological samples a reasonable view is the spatial distribution of the referenced selection.

Anyhow it must be emphasised that the number of possible views is not limited, neither from the knowledge base nor from the implementation. Spatial and cartographic methods provide only a very restricted tool set for supporting sciences for their complex tasks. For example, more complex examples from the same context could use more advanced presentation methods than available from spatial procedures. As it is obvious from this, the implementation of the knowledge resources architecture can be used for any purpose.

With the suggested workflow, the objects from the knowledge resources can be processed by any means like phonetic search, e.g., via classical or modified Soundex algorithms.

 ISBN: 978-1-61208-251-6

This includes the flexible development of a non-limited number of extensions for dynamical search and analysis. It, too, provides a multiplicity of granularity regarding objects and classification.

Any features and data shown, based on the knowledge resources [1] and further sources, even if much less structured like online encyclopedia material, are resulting from the request and workflow, e.g., selection of classification, topics, object, secondary data, area, map projection, applications and so on [9], [10], [11].

One possible example of an algorithm for the interface workflow, with one request iteration is: Knowledge base request, keyword filtering, object processing, UDC filtering, object element processing, object container retrieval, media retrieval, media and container processing, building resulting media, visualisation, provisioning results. This can be used to create multi-disciplinary text and media results, e.g., dynamical distribution maps, from requests, using calculation, processing, and computation of objects. A workflow can enter the knowledge matrix from different directions, e.g., from topic to related topic or from overview to detailed view as well as vice versa.

## VII. Evaluation

The integration of structure and classification allows to use the benefits of algorithms like filtering for any possible use of processing and computing. Structuring the content and context documentation allows a flexible balance for redundancy of data and compute requirements for various application scenarios, even with identical data.

The facetted classification and multi-disciplinary data have proved to provide significant benefits for knowledge reuse and discovery. This includes various ways of describing aspects correctly. From one view a glass of water is half full. From another view the same glass of water is half empty. The two groups representing the classical views might argue that the other view is unintelligible. Both are generally not good as they only represent views. An alternative view will be describing the status giving a filling percentage. In addition this reduces the limitation of unprecise references.

Most content, tasks, and developments handled with information and computing systems are not suitable for any long-term use. The use of the universal knowledge resources and collaboration framework has shown to be very flexible and extendable with implementations and technologies over several decades.

It has been found that standard search and pattern recognition on information is by far not sufficient to gain reasonable results for long-term knowledge herding and evaluation processes. In contrast, the implementation shows excellent results with opening multi-discipline data for IICS, advanced computing, and processing. Statistically, filtering 1 GB of unstructured data delivers less quality than using 10 MB structured classified knowledge base data. The Quality of Data (QoD) must be drastically improved in order to get better results. This can help to reduce compute times, storage volume, and besides overall costs it can help to decrease energy consumption in the end. Using UDC in this context, the availability of a full UDC catalog, and an implementation allowing classification views, combined classification, and ranking priority has proven to drastically increase the QoD. With multi-disciplinary networks, there is even need for a tolerance of individual classification.

In common environments it is only feasible to do one implementation for a specific application, as has been done with these components. Anyhow it has been possible to implement the applications on various architectures providing different resources. Workflows support the use of remote resources (Table IV). In case of a 1000 knowledge-objects reference chain, with 1–10 elements per object, performance will increase much with low latencies.

Table IV
Workflow processes (remote, Ethernet, 1000 nodes).

| *Remote Workflow Process* | *Elements* | *Response Time* |
|---|---|---|
| Knowledge base request | 1000 | 5 s |
| Processing (object, media) | 10 | 7 s |
| Building result | 10 | 5 s |
| Visualisation | 2 | 25 s |

When using one of the described very basic application scenarios on a certain resources architecture the efficiency mostly depends on the decision for the depth of the cross-links to be considered and on the processing requirements for the media data for the originary resources. With current sizes for digital photos and a low depth of five to ten for the cross-links a medium sized application can easily use about one-hundred parallel processes. On a common compute resource without a queue configured for the jobs the response time will be less than a minute.

That way, implementing components with IICS on many compute nodes can profit from using various technologies as suited for different purposes, using task and thread parallelism to the extend needed to handle a problem remotely: *High level*: Integrated Systems, collaboration frameworks, Partitioned Global Address Space (PGAS) models. *Virtualisation level*: Parallel Virtual Machine (PVM). *Low level*: Message Passing Interface (MPI) and OpenMP.

## VIII. Conclusion and future work

It has been demonstrated how complex systems for multi-disciplinary documentation and computing can be built, developed, and extended based on creating long-term-knowledge resources supported by a universal classification and implementing IICS systems. This paper presented the successful implementation of a new universal framework for an integrated system, integrating knowledge resources and implementation components for long-term knowledge

 ISBN: 978-1-61208-251-6 

creation and use, including the facilities for High End Computing and processing resources.

The geoscientific and archaeological knowledge resources have been, structured, extended, and developed for several decades now, having been successfully used with various technology over time. Huge benefits creating new instances of objects and components result from enabling a long-term stepwise development for all parts of the knowledge and application space and a free extendability of the knowledge base. The previous work this implementation is built on has been discussed.

The architecture allows any kind of documentation and algorithm for content, context, information and resources usage. The services and resources usage is very economic and only limited by the limiting implementation factors, e.g., capacities and policies. This solution goes far beyond data and text mining or image analysis and pattern recognition. As shown, classification, as well as spatial data should be integrated with the objects. In no case is it suitable regarding the long-term goals of knowledge creation to "fix" knowledge objects with an application or implementation, neither simple or complex, nor closed or open licensed.

The comparison showed that the possibility of combining methods (UDC, keyword, full-text analysis) does lead to unique benefits. Comparable precision, reliability, performance, and scalability is not available from any isolated method. For any advanced knowledge resources and improved QoD, a flexible classification is undispensable. Bringing the integration of universal classification and IICS into wider acceptance can provide a time-capsule against the transience of knowledge and open new synergetic long-term possibilities.

Complex systems can be created and extended over the necessarily long periods of time, using IICS and UDC. Advanced scientific computing is supported by interfaces, accessing compute and storage resources. Further it has been shown what information and knowledge on content and context can be preserved for medium- and long-term usage even for large complexity of an overall system.

The basic architecture has been presented using a long-term knowledge base (LX), documentation, and classification of objects, the "Collaboration house" framework, flexible algorithms, workflows and dynamical and Active Source components for creating future IICS. Besides that, there is a strong demand for future education and teaching in all disciplines of academia and research in order to mediate and disseminate the basics of knowledge creation and classification.

## ACKNOWLEDGEMENTS

We are grateful to all national and international partners in the GEXI cooperations for the innovative constructive work and to the scientific colleagues at the Leibniz Universität Hannover, the Institute for Legal Informatics (IRI), and WWU Münster, sharing experiences on ZIV, HLRN, Grid, and Cloud resources and for participating in fruitful case studies as well as the participants of the EULISP Programme for prolific scientific discussion over the last years. We do thank the international colleagues from geosciences, informatics, and archaeology in the present collaborations for constructive feedback and proof-reading this paper.

## REFERENCES

[1] LX-Project, "LX Foundation Scientific Resources," 2012, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/#LX (Information) [accessed: 2012-09-29].

[2] C.-P. Rückemann, "Integrating Information Systems and Scientific Computing," *Int. Journal on Advances in Systems and Measurements*, 2012, ISSN: 1942-261x, (to appear).

[3] "Geo Exploration and Information (GEXI)," 1996, 1999, 2010, 2012, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/index.html#GEXI (Information) [accessed: 2012-09-29].

[4] C.-P. Rückemann, "Beitrag zur Realisierung portabler Komponenten für Geoinformationssysteme. Ein Konzept zur ereignisgesteuerten und dynamischen Visualisierung und Aufbereitung geowissenschaftlicher Daten," Diss., WWU, Münster, Deutschland, 2001, 161 (xxii+139) S., URL: http://www.user.uni-hannover.de/cpr/x/publ/2001/dissertation/wwwmath.uni-muenster.de/cs/u/ruckema/x/dis/download/dis3acro.pdf [accessed: 2012-09-11].

[5] C.-P. Rückemann, "Enabling Dynamical Use of Integrated Systems and Scientific Supercomputing Resources for Archaeological Information Systems," in *Proceedings of the International Conference on Advanced Communications and Computation (INFOCOMP 2012), October 21–26, 2012, Venice, Italy*. XPS, Xpert Publishing Services, 2012, pp. 36–41, Rückemann, C.-P., Dini, P., Hommel, W., Pankowska, M., and Schubert, L. (eds.), ISBN-13: 978-1-61208-226-4, URL: http://www.thinkmind.org/download.php?articleid=infocomp_2012_3_10_10012 [accessed: 2012-11-18].

[6] Universal Decimal Classification Consortium (UDCC), "Universal Decimal Classification (UDC)," 2012, URL: http://www.udcc.org [accessed: 2012-02-19].

[7] "Universal Decimal Classification (UDC)," 2012, Wikipedia, URL: http://en.wikipedia.org/wiki/Universal_Decimal_Classification [accessed: 2012-09-09].

[8] B. A. Worley, "Knowledge Discovery from Data," *International Panel on Future High End Systems: Chances and Challenges for Intelligent Applications and Infrastructures, October 22, 2012, The International Conference on Advanced Communications and Computation (INFOCOMP 2012), October 21–26, 2012, Venice, Italy*, 2012, URL: http://www.iaria.org/conferences2012/ProgramINFOCOMP12.html [accessed: 2012-11-17].

[9] B. Steinberger, "Plumes in a convecting mantle: Models and observations for individual hotspots," *JGR*, vol. 105, pp. 11 127–11 152, 2000.

[10] "GMT - Generic Mapping Tools," 2012, URL: http://imina.soest.hawaii.edu/gmt [accessed: 2012-09-02].

[11] GDAL Development Team, *GDAL - Geospatial Data Abstraction Library*, Open Source Geospatial Foundation, 2012, URL: http://www.gdal.org [accessed: 2012-07-15].

 ISBN: 978-1-61208-251-6 

# Marshaling and Un-marshaling CityGML Using Various XML Bindings Techniques

Peter Follo, Robert Forsgren and Gustav Tolt
Sensor Informatics unit
FOI (Swedish Defence Research Agency)
Linköping, Sweden
peter.follo@foi.se, robert.forsgren@foi.se, gustav.tolt@foi.se

***Abstract*— This paper presents the methods and initial results of a study at the Swedish Defence Research Agency in 2012, aiming at investigating various techniques for binding a complex XML schema, such as the OGC's CityGML schema, to a database. This form of binding technique performs a mapping between XML objects and database objects thereby enabling the use of relational databases for update and input/output of customized CityGML models. The results so far indicate that this might be possible with today's framework but not without an undesirable work effort. Also, the lack of documentation makes this work even harder. Follow-on work is planned for 2013 and will provide a basis for an assessment of the frameworks. In this paper, we propose a method to evaluate such frameworks.**

***Keywords-CityGML; XML binding; marshaling; un-marshaling***

## I. INTRODUCTION

The ever-increasing amount of geospatial data and the demand to exchange data within and between authorities and companies drive the demand of standards and regulated ways to fulfil this. One initiative to achieve this is the EU directive INSPIRE [1] stating that all geospatial data that affect environmental domains should be accessible to all member states within the EU. By 2018, these data should even be accessible online. On the national level, in Sweden, the answer to this directive is the implementation of a geodata portal lead by the Swedish mapping, cadastral and land registration authority [2]. Furthermore, many commercial initiatives are possible only if governmental geospatial data are freely available online [3]. This leads to more standards with increasing complexity being developed. One such example is the CityGML XML schema [4] from the Open Geospatial Consortium (OGC) [5].

The question, then, arises how to handle vast amounts of complex data – how to store and maintain it, how to keep up with emerging standards and how to prevent vendor lock-in? Digital geographical data, especially vector layers, are commonly stored in relational databases. It is not to stretch a thought to its limit to believe that also much of these emerging, more complex data structures will be, or at least are preferred to be, stored in relational databases, at least from the point of view of municipal authorities. One example of this is the 3DCityDB [6] developed at Technische Universität Berlin, which stores CityGML models in an Oracle Spatial database. While complex data can be represented in many ways, it is believed that one of the most common ways will be in XML documents with a supporting schema, which is why this paper focuses on such. Even though some relational databases, i.e., Oracle Spatial [7], handle XML directly, this is often done with a severe performance penalty and decreased functionality [8][9] and thus the need to handle complex data structures in traditional database schemes arises. Pure XML databases exist [10]; but, they often lack the GIS functionality present in other more traditional databases such as Oracle Spatial and PostgreSQL /PostGIS [11]. XML databases are also rare in the municipal community where often significant investments in time, money and education related to relational databases have already been made. Several frameworks for (semi-)automatically mapping XML schemes to data structures exist, but not many highlight how to deal with database storage.

This paper proposes a method for evaluating such frameworks and studies a selection of them for their potential ability to marshal and un-marshal CityGML models from and to a database. Since these frameworks are complex and often consist of frameworks-of-frameworks one needs a structured and efficient way to evaluate them. The paper also highlights some of the difficulties encountered with this semi-automatic approach. Our study focuses on the process of creating data structures to create and modify CityGML objects. This is done through automatic generation of source code and libraries from the CityGML XML schema. The goal is to achieve this with a reasonable work effort. CityGML references more than 40 other XML schemes making per schema adoption a cumbersome exercise. Some of the techniques studied in this paper may fail not because of technical impossibilities, but rather on impractical workload or lack of documentation.

Three XML schemas to data objects frameworks have been evaluated: Oracle's Java XML bindings (JAXB), Castor and Apache's Xmlbeans. Two data objects to Data Definition Language (DDL) frameworks, Castor and Hyperjaxb3 (the later based on JAXB and the Java Persistence Application Programming Interface (JPA)) were studied to generate DDL scripts. All these frameworks generate Java sources and the ones capable of generating DDL scripts are highly configurable in choice of database.

 ISBN: 978-1-61208-251-6 

Other framework exists such as Apache's JaxMe, Extensible Stylesheet Language Transformations (XSLT) and others. Time, resources and in some cases technical difficulties prevented these and others to be evaluated in this study.

The remainder of this paper is organized as follows. The next section presents the method and the setup for the experiments conducted. Then, the Experiments section presents the conducted experiments and some intermediate findings. In Section IV, the results are summarized and figures of merits from the experiments are shown. The last chapter, Conclusions, presents the authors' beliefs regarding the feasibility to proceed further with the described method.

## II. METHOD

The proposed method consists of three phases with increasing complexity.

The first phase, Object Model Generation, examines the framework's ability to generate compilable and useful source code. Within this phase we start with a minimal set of customizations and later on add more customizations when needed. When the source finally is generated the compilation stage starts. Some frameworks offer to do this automatically. If this fails we import the generated source code into some Integrated Development Environment (IDE) such as Netbeans and try to compile it that way. On success, to prove usability, we un-marshal and marshal one or more documents and compares the result with the original.

If the first phase is successful it is time to focus on the framework's DDL abilities which are done in the second phase, Data Model Generation. Again, with a similar approach as in phase 1, we try to generate DDL scripts with a minimal set of customizations. It is anticipated that this phase will generate very complex, maybe even impractical, data models, and these generated DDL scripts should not to be considered as the final product. If possible though; un-marshaling and marshaling should be done to prove no loss of data.

The third and last phase, Application Customization, focuses on the framework's abilities to customize the data model to the user's need and the database's capabilities, such as the ability to handle geometries and coordinate transformations. In addition, many times it is not necessary to store every part of the XML structure in a structured way. To reduce the complexity of the data model one might have to simplify the model. The third phase investigates if this is possible and what implications this has on reduced functionality and information loss.

## III. EXPERIMENTS

In this section, we describe the work conducted for the first two experimental phases along with observations made during the work. The experiments are described in a chronological manner reflecting the difficulties that could arise along the way.

To evaluate the frameworks a LOD 3 sample data set from citygml.org covering a street in Frankfurt was downloaded and used. Each experiment was divided in three phases, as stipulated by the proposed method; however no framework reached the third phase due to technical difficulties.

The following (semi-)automatic XML bindings frameworks was used:

- JAXB version 2.2.4 (from the JDK)
- Hyperjaxb3 version 0.5.6
- Castor version 1.3
- Apache Xmlbeans version 2.4.0-7.

Among these frameworks, only Castor and Hyperjaxb3 have the ability to generate DDL scripts.

All experiments were conducted on a Dell XPS M1330 laptop with an Intel Dual Core 2.2 GHz Processor (T7500) and 4 GB of RAM. Operating system of choice was Fedora 16. Java version was Oracle's JDK version 1.7.0_04. Netbeans 7.0.1 was used as Java IDE.

### A. Phase 1: Object Model Generation

#### 1) Oracle JAXB 2.2.4

First try: generate and build a Java library using JAXB's *xjc* without any customization. This yielded name conflicts in the included XML schemes due to multiple definitions of elements with the same name. Some of them should really not be a conflict since they should be defined in different namespaces, i.e., the element *Role* defined both in cityObjectGroup.xsd and xlink.xsd. *xjc* is kind enough to output some hints to overcome the name clashes. Thus, the second try was to generate and build a Java library with a minimal set of customizations. Several conflicts arise due to XML names being converted to the same Java name, i.e., the XML name *_Solid* transforms to the Java name *Solid* which conflicts with a previous defined XML name *Solid* also transformed to the Java name *Solid*. Thirteen customizations were necessary due to conflicts in the XML schemas. Thereafter, the errors change focus to conflicts within the package to be built, conflicts sometimes already addressed in earlier customization, some of them really not conflicts at all. The third try was then to generate sources and build in a separate step. 594 Java source files were generated in about 10 seconds, including the time for internet access to the referenced schemes. The sources were imported into Netbeans and compiled without errors in 12 seconds. Marshaling and un-marshaling of the sample data set went without errors. Visual random inspection of the original data and the marshaled data indicated no errors or data loss.

#### 2) Apache Xmlbeans 2.4.0-7

First try: generate and build Java library without any customizations. The *SchemaCompiler* worked out of the box generating 2526 classes in 6 seconds and building a Java library in 55 seconds with one ignorable warning about classpath settings. Marshaling and un-marshaling went without problems.

#### 3) Castor 1.3

First try: generate and build Java libraries without any customizations. This resulted in a null pointer exception complaining on missing parent for "the built in parent type for: MeasureOrNullListType". No sources were generated. *MeasureOrNullListType* origins from the referenced GML 3.1.1 schema basicTypes.xsd.

 ISBN: 978-1-61208-251-6 

Second try: generate sources for basicTypes.xsd. This yielded in more errors of the same type, this time pointing to *CountExtentType*. Some sources where generated though. No workaround has been found not implying much work or unacceptable loss of functionality why castor sadly has to leave the stage. A bug report (CASTOR-3223) has been filed.

### B. Phase 2: Data Model Generation

#### 1) Oracle JAXB 2.2.4/Hyperjaxb3 0.5.6

The first try was to examine the possibilities to add JPA annotations to the generated classes. The goal is to annotate every *complexType* as an Entity class. This was possible with the JAXB Annotate plugin [12], but only feasible if enabling the undocumented feature to allow multiple matches from the XPath expressions. Otherwise, one has to do a customization for each and every of the generated classes. Moreover, the JAXB Annotate plugin extensions were not allowed in the global binding scope, leading to a schema binding for each and every imported and included schema. However, Hyperjaxb3 targets this specific question, being a plugin to JAXB linking JAXB to JPA. Hyperjaxb3 also takes into account many other problems that arise when bridging the object model and data model [13]. Thus, the second try was to do the same experiment with Hyperjaxb3: Close to twenty customizations were made and given to Hyperjaxb3. 770 classes were generated, but the generated code did not compile due to invalid arguments to some methods. This occurs in four of the generated classes but it is manageable by manually editing the files. After editing, the process of generating DDL scripts starts but soon bails out again. According to the generated output, it is suggested that the hibernate *EntityManager* jars are missing. Further investigation could not conclude that this is the case; the needed jars seems to be included.

## IV. RESULTS

Table 1 summarizes some figure of merits for the different frameworks together with some important notes. Oracle's JAXB 2.2.4 passes phase 1 but must be paired with some other technique to reach phase 3 of the experiments. This could be done with standalone JPA annotations but since hyperjaxb3 exists which do exactly this and more this was not further investigated in this study. Hyperjaxb3 was the most successful framework reaching the DDL generation phase but fails on a configuration error. No work-around was found. Hyperjaxb3 is a maven plugin, using several other artefacts. It is hard to follow exactly where or in what artefact it goes wrong. Also, it is hard to understand what is configurable or not from hyperjaxb3's point of view. Some configurations applied in the pom.xml file for the underlying artefacts did not follow through. Documentation to clarify this is desirable. Castor 1.3 fails with a null pointer exception not finishing phase 1. No documentation was found indicating what went wrong. To understand why one has to dig into the source code but this is out of scope of this study. Apache Xmlbeans 2.4.0-7 went without errors. No customizations was needed what so ever and marshalling and un-marshalling went without errors. Unfortunately, Xmlbeans is not able to generate DDL scripts.

TABLE I. FIGURE OF MERITS FOR THE DIFFERENT FRAMEWORKS.

| Framework | Cust. | Classes | Note |
|---|---|---|---|
| Oracle's JAXB 2.2.4 | 3 | 594 | Passes phase 1. Not applicable to phase 2 & 3 standalone. |
| Hyperjaxb3 0.5.6 | 16 | 770 | Passes phase 1 but fails in phase 2 during DDL generation. Generated code does not compile without editing. Build must be done in a separate step. |
| Castor 1.3 | 7 | - | Fails on phase 1. |
| Apache Xmlbeans 2.4.0-7 | 0 | 2526 | Passes phase 1. No customizations needed. Not applicable to phase 2 & 3 standalone. |

## V. CONCLUSIONS

While promising, these frameworks still seem far from being capable to handle large complex XML schemas such as OGC's CityGML schema out-of-the-box. If not by technical means, so by the lack of documentation. The lack of documentation and the fact that these frameworks often are frameworks-of-frameworks makes debugging and understanding of the internal processes hard. Yes, the source code is there, but with productive aspects from e.g. a municipal agency's point of view it is not realistic to dig that deep into a problem. First and foremost, these frameworks must be better documented, to make it possible to know if you are trying to solve your problem the right way. This will give the developers the proper feedback to make the frameworks easier to use and in the long term more robust. Better documentation will also ease the burden of the developers and the community to answer newbie questions and it will also ease the burden of the users to adopt these techniques. In the (failed) experiments it is pointed out, several times, that no work-around was found, although this does not mean that such does not exist.

## VI. FUTURE WORK

In this paper, we have investigated three frameworks for binding a complex XML schema to a database.

One of the most promising, due to its ease-of-use, frameworks is Apache Xmlbeans. However, it could not be investigated thoroughly enough up to this point due to time and resource constraints, but a deeper analysis is scheduled as an upcoming action. We also note that this framework most be combined with other techniques to reach the third phase of the experiments.

From what we have experienced so far, Hyperjaxb3 should also be further investigated to see whether it is possible to circumvent the configuration error described in Section III.B.1. It must be investigated how the hyperjaxb3 maven plugin interacts with other maven plugins, especially

 ISBN: 978-1-61208-251-6 

the *hibernate* and *ant task* plugins. Lack of time prevented us from addressing this issue thoroughly in this study, but will instead be considered as a next step.

It is also desirable to dig deeper to find the exact reason for processing errors, and for this we anticipate that working closer together with the developers and the community is a necessity.

A better XML comparison technique must be developed; we must be able to compare large XML files in a more clever way than manual inspection to guarantee integrity of the object during marshaling and un-marshaling; taking into account the fact that the object can be represented in many valid ways in an XML structure.

## ACKNOWLEDGMENTS

Thanks to the developers of these frameworks, working under high pressure and short of resources, sometimes sacrificing their spare time to answer questions and fix bugs. Without your efforts this would be a daunting task, maybe even impossible. Thanks also to the companies sponsoring them in their work working along with the open source community.

## REFERENCES

[1] EU, "Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)", European Commission, 2007

[2] Lantmäteriet - the Swedish mapping, cadastral and land registration authority, "Geodataportalen", Accessed Dec 2012: http://www.geodata.se/GeodataExplorer/start.jsp?loc=en

[3] G. Vickery, "Review of recent studies on PSI re-use and related market developments", European Commission, 2011

[4] Open Geospatial Consortium, "OpenGIS® City Geography Markup Language (CityGML) Encoding Standard", OGC, 2008

[5] Open Geospatial Consortium, "About OGC|OGC®", Accessed Jan 2013: http://www.opengeospatial.org/ogc

[6] A. Stadler, C. Nagel, G. König. and T.H. Kolbe, "Making interoperability persistent: A 3D geo database based on CityGML", Lee & Zlatanova (eds.), 3D Geo-Information Sciences, Selected papers from the 3rd International Workshop on 3D Geo-Information, Seoul, Korea. LNG&C series, Springer Verlag, pp. 175-192. 2008

[7] Oracle, "Oracle Database Online Documentation", Accessed Aug 2012: http://www.oracle.com/pls/db111/portal.portal_db?frame=&selected=7

[8] PostgreSQL, "PostgreSQL: Manuals", Accessed Dec 2012: http://www.postgresql.org/docs/manuals/

[9] MySQL, "MySQL :: MySQL 5.5 Reference manual", Accessed Dec 2012, http://dev.mysql.com/doc/refman/5.5/en/index.html

[10] "XML database", Wikipedia, Accessed Dec 2012: http://en.wikipedia.org/wiki/XML_database

[11] PostGIS, "PostGIS: Home", Accessed Dec 2012: http://www.postgis.org/

[12] A. Valikov, "Annotate Plugin - Confluence", Accessed Dec 2012: http://confluence.highsource.org/display/J2B/Annotate+Plugin

[13] A. Valikov, "JAXB vs. JPA - Confluence", Accessed Dec 2012: http://confluence.highsource.org/display/HJ3/JAXB+vs.+JPA

 ISBN: 978-1-61208-251-6 

# Time Geographic Network Modeling for Restraint Space of Transportation Network

Zhang-cai Yin, Qian-yao Duan, Hua-tao Sun, Qing-quan Liu, Yu Bao, Zhi-qiang Wang
School of Resources and Environmental Engineering
Wuhan University of Technology
Wuhan, China
yinzhangcai@163.com; littexiaobai@qq.com; buotaoxiongyong@foxmail.com; 786931088@qq.com; 982771045@qq.com; 472833494@qq.com

***Abstract*—Time geographic accessibility in heterogeneous space can be converted to the shortest path problem. Based on raster grids, current time geography theory analyzes the shortest path among grid cells by using 8-neighbor mode (network). However, due to the irregularity of transportation network, the network is not constantly included in 8-neighbor network. This means neighbor network paths are different from continuous-space paths in location and length, and as cause a time distance error. In order to address this problem, the paper focuses on the restraint space of irregular transportation network, and introduces Voronoi diagram grid. Then, we establish a neighbor network model based on Delaunay Triangulated Irregular Network to contain transportation network. This means neighbor network paths can converge to continuous-space paths under the same condition of start and end, and as decrease a time distance error. As a result, the shortest path between Thiessen polygons is calculated with less error of time distance. Thus, we have the apporach to model time geographic network for restraint space of transportation network. This article also describes implementation of the models using ArcGIS 9.3 with a real transportation network as a way of validation.**

***Keywords-network mode; irregular triangle network; field; space-time mode; raster***

## I. INTRODUCTION

Time geography's main concern is measuring the accessibility region. Since the gradual space can be divided into raster grids, it enables the grid cell's accessibility problem to be converted to the shortest raster path problem [1]. The raster path consists of a number of adjacent grid cells, the sum of whose weights is the path length. The weight of a cell is set to be the shortest time for passing the cell. Assuming the shortest time for passing a cell $c$ horizontally or vertically is actually $t_h$ and $t_v$ respectively. Without loss of generality, we assume $t_h > t_v$, thus the weight of $c$, $t_c = t_v$; in such a way, $t_c$ would replace $t_h$ to be the shortest time for crossing $c$ horizontally in the shortest path algorithm analysis with a error: $t_h - t_v$. However, the error would be quite small in gradual space, and get larger in transportation network restraint space (TNRS). For example, if $c$ only has one vertical road, the result of $t_h - t_v$ would be quite large. This means we expect time geography in TNRS.

Miller introduced a time geography field for TNRS [2]. It describes how the time geography network models can be established for the city space by adapting raster grid and its 8-neighbor mode or network (neighbor network, NN). In the city, Salt Lake City, urban roads crisscross. A raster unit of 500m×500m either has no road, or has a crisscross network of roads. Thus, $t_h$ and $t_v$ of the unit are pretty much the same, resulting in small error: $t_h - t_v$. According to the trace data of individuals, Miller verified the efficiency of raster time geography field model in regular-TNRS, such as Manhattan, Salt Lake City.

So far, however, the elementary theory of time geography in TNRS deals with regular transportation network only. Since raster grids are regular rectangle units, the subdivision of units can hardly allow for transportation irregularity. Some units would only contain a single route (for instance, the unit $c$ only has a vertical road), which may have larger weight errors. We would study the mathematical basis of time geography in irregular-TNRS. As a result, we will have a complete time geography theory for TNRS.

Heterogeneous space's time geography is an extension of (classical) homogenous space's time geography based on fields. Raster time geography field is regular-based, and this article would introduce irregular Voronoi diagram field, to reflect the additional transportation network by the neighbor networks based on Delaunay Triangulated Irregular Network (D-TIN). Thus, we can measure the shortest path and spatio-temporal accessibility for TNRS.

The article is structured as follows. Section 2 summarizes the existing time geography. Then, we introduce time geography in TNRS (Section 3). Section 4 gives the approach to construct time geography networks for TRNS. The approach will be implemented and tested in ArcGIS in Section 5. Conclusions close the article (Section 6).

## II. TIME GEOGRAPHY

The fundamental problem of time geography is measuring the accessibility region of an agent within the known time $T$. Given the agent's start point $s$ and velocity distribution $v$, if the agent leaves from $s$ and arrive at $p$ within the shortest time $D(s, p)$, $D(s, p) \leqslant T$, $p$ should be included in the reachable region. When the end point $e$ is also known, another condition must be added: the agent can leave from $p$ and reach $e$ within the minimum time interval $D(p, e)$, $D(s, p)+D(p, e) \leqslant T$. In this way, the accessibility of $p$ can be converted to the shortest-path problem [1]. $D(s, p)$ and $D(p, e)$ are corresponding to path $S(s, p)$ and $S(p, e)$ respectively.

 ISBN: 978-1-61208-251-6 

In homogenous space, *v* is everywhere equal, so the shortest paths *S*(*s*, *p*) and *S*(*p*, *e*) are line segments. Classical time geography uses cones to show an agent's reachable area while moving freely at *s* (Fig. 1a), and uses prisms to represent that while proceeding directed movements from *s* to *e* (Fig. 2a) [3].

In heterogeneous space, *v* is not constantly the same, so *S*(*s*, *p*) and *S*(*p*, *e*) may not always be straight, but curves. The algorithms of finding the shortest path curves are almost network-based, such as single-source shortest path algorithm (Dijkstra). This means the accessibility problem, via the shortest-path problem, can be converted to the network-modeling problem, including space grid-modeling and its NN-modeling. Therefore, the process to measure accessibility region is 3 steps:

*1) to construct the grid and its NN;*

*2) to analyze the shortest path S(s, p) and S(p, e);*

*3) to determine whether p is reachable, and to construct the accessible region including reachable point p or the discrete cell whose center is p.*

Transportation network is heterogeneous space, and is a NN covering itself, and thereby only from step 3) Kwan [4] analyzed time-space accessibility, in which the time-space prism degrades into vertical sections based on transportation network [5], shown as Fig. 2b.

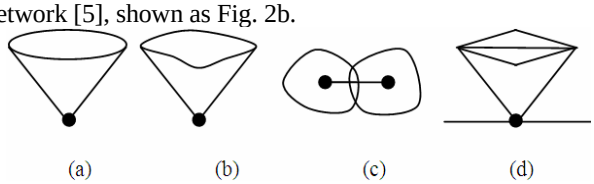



Figure 1. (a) time-space circular cone; (b) time-space cone; (c) intersection of two reachable regions; (d) the time-space cone of transportation network in restraint space

TNRS is also heterogeneous, and can be divided to raster grid [2], by which enabling us to investigate time-space accessibility [6, 7, 8, 9] to analyze and visualize the individual's reachable region in GIS [10]. The research on time geography application is beyond this paper; here, it is important to note that network modeling for TNRS has so far stayed in NN indirectly representing transportation network.

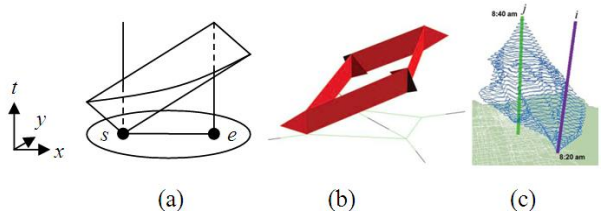



Figure 2. (a) the prism in homogenous space; (b) the prism of transportation network; (c) spatial prism of regular transportation network

It is simple to establish NN indirectly, representing transportation network. After TNRS is transformed to raster grids, we usually turn to an auxiliary NN to investigate the shortest-paths between raster cells [9]. The NN can be constructed by two restraints [6, 8]:

*1) the center of each cell is a network node;*

*2) the lines connecting adjacent centers are the network edges.*

In general, common adjacent relationship includes Rock mode (4-neighbor), Queen mode (8-neighbor) and Knight mode (16-neighbor and its extension), representing adjacent relations among discrete units of varied orders. In this paper, the auxiliary NN are called raster NN (some of them are called virtual network [6], planer network view [7], and implicit network [2]). The raster NN's weights can be dynamically generated by simple liner interpolation of raster cells' weights [8, 11]. For example, if both cells with weights $t_a$ and $t_b$ are horizontally adjacent, the edge connecting both cells has weight $(t_a + t_b)/2$. Since transportation network directly influences its restraint space' cell attributes, it affects the weights of NN. There exists a mapping relationship between transportation network and NN: transportation network → cell → NN. In regular-TNRS (like Fig. 3), the travel time in transportation network, via the weight of raster cell, can be transmitted to that of NN effectively. Armed with such a raster NN, Miller measured the space-time prism of Salt Lake City [2] (see Fig. 2c); of course, space-time cone can also be measured, shown as Fig. 1b. The question whether two agents would meet within *T* is answered by testing whether the agents' time-space volumes intersect (Fig. 1c).


Figure 3. Part of Salt Lake City

However, the way to establish NN directly representing transportation network has not been proposed till now. An exception may be YU et al. [8], who suggest 8-NN should add the edges connecting nonadjacent cells to stand for tunnel, but have not yet developed the algorithm for NN to cover plane transportation networks. Generally speaking, there are larger weight errors in the raster NN covering transportation network. The NN of the raster cells only containing a single route is belong to this category. Its reason is that the travel time in transportation net is transmitted to that of NN, via the weight of raster cell. A naive approach to construct NN is to build the mapping relation between transportation network and raster NN, and thus, transportation net' weights can be directly passed to raster NN rather than in an indirect way bypassing raster cells' weights, without cells' weight errors transmitting to the NN. But this approach is wrong, as we will explain later. Fig.1d describes the spatio-temporal reachable region in a single road restraint space.

 ISBN: 978-1-61208-251-6 

## III. TIME GEOGRAPHY IN TRNS

### A. *Voronoi Diagram and Delaunay Network*

In the realm of GIS, Voronoi diagram is a significant field model. The field is a general conceptual space model constructing continuous spatial change; its logical model is usually described by specific field model of regular rectangles (such as raster), contours and irregular areas like Voronoi diagram and its dual graph D-TIN. The first two have been applied in raster time geography field [2] and isochrones (lines of equal travel time) [12]. In a multitude of geography spatial analysis, field model based on irregular area may be superior, suitable for the practical distribution of data.

There are two networks: raster grids and its NN in raster time geography field. After surface space is transformed to the raster grids, we can construct a NN to analyze the shortest path between raster cells. Similarly, we construct two networks: Voronoi diagram and its D-TIN, for irregular-TNRS (like Fig. 4). Since the nodes and edges of D-TIN stand for Thiessen polygons and their adjacent relationship respectively, D-TIN is the NN where a weight is the shortest time for passing an edge, and this diagram is the field model. Therefore, we can analyze $S(s, p)$ and $S(p, e)$ by D-TIN. Then, if $p$ is accessible, it is accessible Thiessen polygon whose center is $p$.

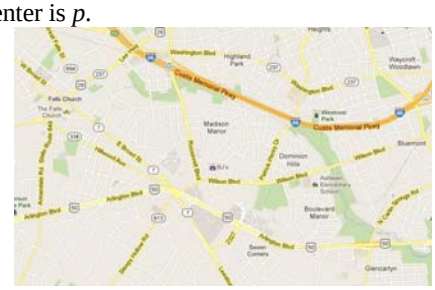


Figure 4. Part of Washington

At present, the ways to model D-TIN for various restraints are fairly mature, such as inserting restraint lines into D-TIN. Theoretically, D-TIN can converge to any transportation network. Consequently, transportation network' weights can be directly passed to D-TIN rather than in an indirect way bypassing Thiessen polygons' weights. That means it can avoid the deviation errors and weight errors, and can decrease the elongation errors, that D-TIN as NN is used to analyze $S(s, p)$ and $S(p, e)$. Below, we describe how to construct D-TIN and D-TIN's superiority to raster NN for irregular-TNRS.

### B. *Time Geography Networks based on Voronoi diagram and D-TIN*

Let us take a crossroad for example (Fig. 5a), to picture the mapping procession from transportation net to NN. 7. Let the point where road AB cuts road CD be called O (Fig. 5a). The weights (the shortest passing time) of AO, BO, CO and DO are all set to be 1. Supposing there are pathways for walk in other edges AD, AC, CB and BD, thus their weights are all set to be 4. As a result, the shortest time from A to B is $D(A, B) = 2$.

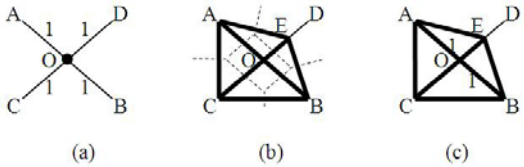



Figure 5. (a) crossroad; (b) Voronoi diagram and its TIN; (c) weighted TIN

Firstly, we picture the mapping from transportation net to D-TIN NN. Armed with the point set {A, B, C, E, O}, D-TIN and Voronoi diagram can be built (Fig. 5b), where E is the midpoint of OD. Again, Supposing there are pathways in edges AE and BE. According to Pythagorean Theorem, the weights of AE and BE are both 3.16. Through transportation' and pathway' net-to-NN maps, we can get the weights of AO, BO and CO with value of 1; that of OE is 1/2. On the other hand, the weights of AC and CB are both 4; that of AE and EB are 3.16. As a result, $D(A, B) = 2$, and $S(A, B) = \{AO, OB\}$, shown as Figure 5c.

Secondly, we picture the mapping from transportation net to raster NN. When this junction O is exactly at the raster corner (Fig. 6a), there are two kinds of projections of transportation network to 4- and 8-raster NN. (1) In 4-NN(Fig. 6b), edges AD, AC, BC and BD represent transportation net's four circuits, AOD, AOC, COB and BOD respectively with weights of 2, so $D(A, B) = 4$. When 4-NN edges AD, AC, BC and BD represent pathways, their weights are 4, and so $D(A, B) = 8$. (2) In 8-NN, edges AB and CD stand for the routes AOB and COD respectively, with weights of 2. However, AB and CD do not intersect at O, shown as Fig. 6c. This means, when raster NN cannot cover the transportation net, the gaps between them will cause differences between transportation net paths and raster NN paths in location and length. This, in turn, introduces elongation errors (NN path being longer than transportation network path) and deviation errors [2] in the shortest path algorithm analysis and accessibility analysis. It answers the question why transportation net' weights cannot be directly passed to raster NN.

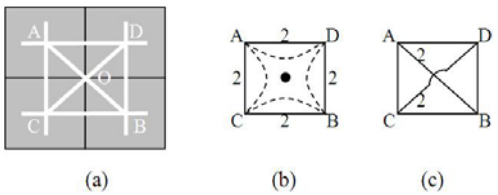



Figure 6. (a) transportation network and raster grid; (b) 4-neighbor network simulating transportation network; (c) 8-neighbor network simulating transportation network

Together, on condition that the transportation network is regular, that is, each cell has crisscross roads or no road, such as Manhattan net, raster NN based on Rock mode exactly seamlessly covers the transportation network, and avoids deviation error and reduces elongation error. The raster time geography field applied to Salt Lake City is belong to this

 ISBN: 978-1-61208-251-6 

category. Also, that indicates the necessity of using NN to directly cover transportation network. However, when raster NN covers a single route in irregular transportation net, it introduces larger weight error; when does not cover the net, the gaps between raster NN and irregular transportation net introduce these errors of elongation, deviation, topology inconsistency. Hence, raster NN is unable to adapt to irregular-TNRS. This means, to enable the NN to cover irregular transportation net, we need introduce a kind of irregular NN and realize the mapping relation between NN and transportation net. The NN illustrating irregular transportation network is also irregular and D-TIN can seamlessly show the transportation network. In sum, the paper studies time geography theory for irregular-TNRS on condition that D-TIN is the prototype of NN and Voronoi dual graph is the field or grid model.

## IV. CONSTRUCTION OF TIME GEOGRAPHY NETWORKS FOR TNRS

After studying on the theoretical issue of modeling NN and grid based on transportation network constraint, we would pay much attention to the mapping from transportation network to D-TIN NN, and to establish Voronoi diagram grid.

### *A. Time geography network modeling mechanics based on transportation network constraint*

NN in TNRS, on the one hand, represents the centers of cells and their adjacent relations, so it is restrained by the spatial grid which subdivides TNRS into cells (grid →NN); on the other, have to contain the transportation net, so it is bound by transportation net (transportation net → NN). Governed by grid, the centroids of discrete units may not lay on the transportation net be located in the corresponding units. Here comes to the problem that how to build the NN on above both conditions. This directly influences the time geography network constructing for TNRS.

Since there generally exists a one-to-one relationship between the grid and its NN, hence, the one-way constraint relation: grid→NN, can be extended to a two-way relation: grid←→NN. This means that NN can also control the grid subdivision for TNRS; that is NN → grid. Due to the objectivity of transportation net, the subjectivity of grid and its NN, the priority of the constraint: transportation net→NN is higher than others, such as NN → grid. It is for such reasons that, we can reduce the double constraints on modeling NN to both relationships: transportation net→NN, and NN→grid, or transportation net→NN→grid (Fig. 7).

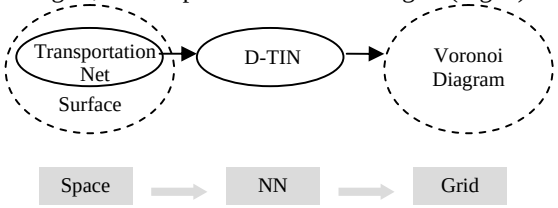



Figure 7. The model of constructing time geography network for TNRS

### *B. The mapping relation between transportation net and NN*

On condition that TNRS is divided into transportation net and the restrained surface, we subdivide transportation net into plane transportation net on the flat surface and the non-planar three-dimensional transportation net on or under the surface, like a tunnel. This article only considers plane transportation net. The mapping from transportation net to NN includes discretizing the plane transportation net and the flat surface, and these discrete points based generating D-TIN, namely NN of TNRS. In this case, NN not only presents Thiessen polygons and their adjacency relations, but also contains transportation net, with the result that such double constraints of transportation net and grid on NN were fulfilled.

The D-TIN NN, allowing for the restraint of discrete points in transportation net, we find it difficult to avoid edges of D-TIN being crossed by transportation lines, which disobey the principle that NN completely contains transportation net. For instance, the D-TIN (Fig. 8b), generated from two discrete points A and B on one transportation line (Fig. 8a) and the surrounding points C and D, is crossed by the routine AB (Fig. 8c). If the transportation line is presented by TIN path, (e.g. AC-CB stands for AB), it would deviate the transportation line and cause related errors.

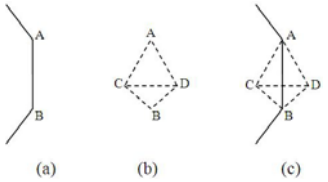



Figure 8. Transportation line crosses TIN (a) transportation routine; (b) TIN; (c) TIN edge intersecting transportation routine

As a result, constructing the D-TIN allowing for transportation line restraint is the key to construct NN. By adding new points, we may re-build D-TIN whose edges would not cross transportation lines. Thus, NN may completely contain transportation net.

### *C. The mapping relation between NN and grid*

According to the theory of dual graph, D-TIN can be converted to the only Voronoi diagram. Since the centroid set of Voronoi diagram namely is the node set of D-TIN, which includes the discrete point set in transportation net. This means the discrete point set is contained in the centroid set. Because a Thiessen polygon has one and only one centroid, the centroid is a discrete point in transportation net, yes or no, Therefore, there is at most a transportation net' discrete point in a Thiessen polygon, and the discrete point must coincide with the centroid of the Thiessen polygon. As a result, the Voronoi diagram is the kind of space subdivision with constraint of transportation net, and it may become a kind of field model in TNRS. Hence, Voronoi diagram field cannot be replaced by raster field regardless of constraint of transportation net, in time geography for TNRS.

 ISBN: 978-1-61208-251-6 

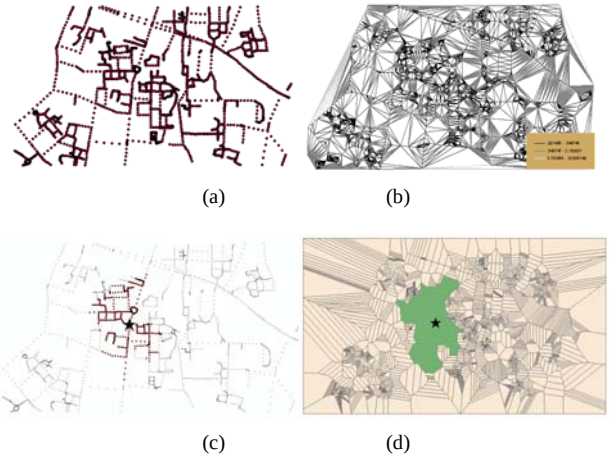
### D. *Calculation of NN's weights*

To put it simple, we assume that the vector $v_1$ is constant in transportation net, which is the same to vector $v_2$ out of transportation net and $v_2 < v_1$. Thus, the weight of a NN edge representing transportation routine can be obtained by dividing routine length by $v_1$; as for a NN edge outside of transportation net, we gain its weight by dividing its length by $v_2$. The nodes and edges representing transportation net in NN should be in correspondence with the real transportation network in node location, topological relation and weight; so it avoids the location deviation error, topological inconsistency error and path elongation error. Besides, since the weight calculations of transportation network and non-transportation network part in NN are relatively isolated, we can avoid weight error of non-transportation network caused by transportation network in NN.

## V. EXAMPLE

This section illustrates the methodology introduced above. For this purpose, the methodology was implemented in ArcGIS 9.3. All figures in this section are computed with ArcGIS 9.3.

Taking part area of WuHan in China for example, we analyze the accessible domain in TNRS without considering 3-D transportation net such as tunnels and bridges. Firstly, according to the discrete points set of transportation net (Fig.9a), we can model the NN based on D-TIN for TNRS, and make sure that NN contains the transportation net completely. Let us suppose that $v_2$ is 60km/h, and $v_1$ is 10km/h. Thus, the weights of NN can be calculated, shown as Fig.9b.



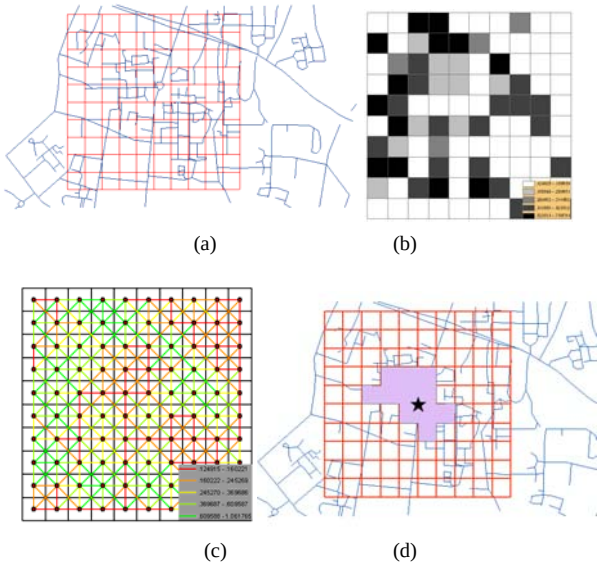
(a) (b)

(c) (d)

Figure 9. (a) discrete points; (b) D-TIN NN; (c) accessible points in NN; (d) accessible domain based on Thiessen polygons

Secondly, we construct Voronoi diagram field based on the discrete points set of transportation net. Finally, the accessible points in NN can be computed in known start point $s$ and $T$ = 30 seconds (Fig.9c), and these points Thiessen polygons comprise the accessible region (Fig.9d).

Similarly, using the principle of time geography field, we can also model the raster grid (Fig.10a), and can analyze the weight of every cell (Fig.10b). After obtaining the weights of 8-NN (Fig.10c), we may measure the accessible points in NN, and their cells which form the accessible region (Fig.10d) under the same conditions of $s$, $T$, $v_1$ and $v_2$.



(a) (b)

(c) (d)

Figure 10. (a) raster gird; (b) gird' weights; (c) 8-NN with weights; (d) accessible domain based on raster cells

Consequently, there are two kinds of different accessible domains for the same TNRS (Fig.11), where the solid line boundary is based on Theissen polygons, and the dotted line boundary is based on raster cells.

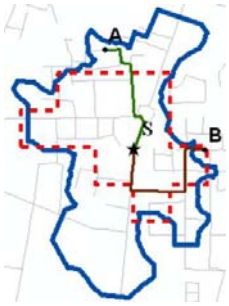



Figure 11. Two kinds of different accessible domains

As shown in the figure, two points A，B are located on the transportation net. It is easy for us to find the shortest path from the start $s$ to A in transportation net, by the Dijkstra algorithm in ArcGIS. Obviously, the path lies also in transportation net, and its shortest passing time is 26.538 seconds. That means an agent can arrive at A from $s$ within the given time $T$. Likewise, the shortest path from the start $s$ to B, lies in transportation net, and its shortest passing time is 32.316 seconds, so in $T$ the agent cannot arrive at B from $s$. Point A is contained within solid line boundary, and is out of dotted line boundary. On the contrary, point B is out of solid line boundary and is in dotted line boundary.

 ISBN: 978-1-61208-251-6 

## VI. CONCLUSIONS AND FUTURE WORK

Based on the summary of time geography theory such as that of raster time geography field, the article resolves spatio-temporal accessibility into time geography network modeling. In order to model time geographic network for TNRS, this paper introduces Voronoi diagram grid and NN based on D-TIN. After analyzing the relationship among transportation net, grid and NN, the paper approves a modeling strategy that transportation net is contained within NN, while NN controls grid. Finally, on the basis of NN weight calculation, we use instances to analyze time geography network modeling algorithm for TNRS. Future studies mainly consider about non-planar transportation net, anisotropy of transportation network, etc. It is necessary to validate the proposed solution in the paper in the large transportation net.

## ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (Grant No. 41071283), and supported by "the Fundamental Research Funds for the Central Universities" (Grant No. 2012-IV-097).

## REFERENCES

[1] M. F. Goodchild, "An evaluation of lattice solutions to the problem of corridor location". Environment and Planning A, vol. 9, Jul. 1977, pp. 727–38, doi:10.1068/a090727.

[2] H. J. Miller and S. A. Bridwell, "A Field-Based Theory for Time Geography", Annals of the Association of American Geographers, vol. 99, Jan. 2009, pp. 49-75, DOI:10.1080/00045600802471049.

[3] S. Winter and Z.-C. Yin, "Directed movements in probabilistic time geography", International Journal of Geographical Information Science, vol. 24, Sep. 2010, pp. 1349–1365, DOI: 10.1080/13658811003619150.

[4] M. P. Kwan, "Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework", Geographical Analysis, vol. 30, Jul. 1998, pp. 191-216, DOI: 10.1111/j.1538-4632.1998.tb00396.x.

[5] B. Kuijpers and W. Othman, "Modeling uncertainty of moving objects on road networks via space–time prisms", International Journal of Geographical Information Science, vol. 23, Sep. 2009, pp. 1095–1117, DOI:10.1080/13658810802097485.

[6] J. Xu and R. G. Lathrop, "Improving simulation accuracy of spread phenomena in a raster-based geographic information system". International Journal of Geographical Information Systems, vol. 9, Feb. 1995, pp. 153–68, DOI:10.1080/02693799508902031.

[7] W. Collischonn and J. V. Pilar, "A direction dependent least cost path algorithm for roads and canals". International Journal of Geographical Information Science, vol. 14, Aug. 2000, pp. 397–406, DOI:10.1080/13658810050024304.

[8] C. Yu, J. Lee and M. J. Munro-Stasiuk, "Extensions to least-cost path algorithms for roadway planning". International Journal of Geographical Information Science, vol. 17, Nov. 2003, pp. 361–376, DOI:10.1080/1365881031000072645.

[9] B. G. Alexandre, "An extension of GIS-based least-cost path modeling to the location of wide paths", International Journal of Geographical Information Science, vol. 24, Apr. 2010, pp. 983-996, DOI:10.1080/13658810903401016.

[10] X. Chen and M. P. Kwan, "Choice set formation with multiple flexible activities under space-time constraints". International Journal of Geographical Information Science, vol. 26, May. 2012, pp. 941-961, DOI: 10.1080/13658816.2011.624520.

[11] A. K. Saha, M. K. Arora, R. P. Gupta, M. L. Virdi and E. Csaplovics, "GIS-based route planning in landslide-prone areas", International Journal of Geographical Information Science, vol. 19, Feb. 2005, pp. 1149-1175, DOI:10.1080/13658810500105887.

[12] D. O'Sullivan, A. Morrison, J. Shearer, "Using desktop GIS for the investigation of accessibility by public transport: An isochrone approach". International Journal of Geographical Information Science, vol. 14, Aug. 2000, pp. 85–104, DOI:10.1080/136588100240976.

 ISBN: 978-1-61208-251-6 

# Towards a Semantic Interpretation of Satellite Images by Using Spatial Relations Defined in Geographic Standards

Samuel Andrés
*IRD, UMR Espace-Dev*
*Montpellier, France*
*samuel.andres@ird.fr*

Christelle Pierkot
*IRD, UMR Espace-Dev*
*Montpellier, France*
*christelle@pierkot.fr*

Damien Arvor
*IRD, UMR Espace-Dev*
*Montpellier, France*
*damien.arvor@ird.fr*

***Abstract*—To work more efficiently, remote sensing applications need to introduce semantic tools which are able to express experts knowledge. One important part is about spatio-temporally which is often used by experts to make their interpretations. We argue that geographic standards could help to formalize spatio-temporal expertise. In this paper, we focus on spatial relations by giving an example on how to fin beaches by using spatio-temporal ontology.**

***Keywords*-Remote sensing; Semantic interpretation; Geographic standards; Spatial relations.**

## I. INTRODUCTION

Usually, to interpret and recognize objects on satellite images, experts use both remote sensing knowledge (e.g., spectral signature, remote sensing index, etc) which are linked to physical properties of objects (wavelength, texture, etc), and field knowledge (features properties, spatial relations, etc), linked to the application domain such as agriculture, ecology or urban planning. For example, to find *mangrove* in an image of French Guiana, the expert uses knowledge acquired from his field experience (e.g. mangrove grows in salt water, between ocean and continent etc) together with knowledge stemming from remote sensing processes (e.g. NDVI index for vegetation segments, NDWI to find water ). However, this way of proceeding often brings some problems. Indeed, because of this knowledge depends on the expert personal experience, it is diverse and non equivalent. First, two experts do not have the same interpretation of one image because they do not use necessarily the same treatments in the segmentation or classification processes. For example, many indexes exist to classify vegetation (e.g., NDVI, VGI, etc) and the expert will probably use the one he has the best knowledge, thus the selected index for this process is not necessarily the same from one expert to another. Moreover, the expert often proceeds by trial and error tests before finding a suitable solution. Finally, a remote sensing expert has not necessarily the sufficient knowledge of the field to well interpret the image in a specific context. Therefore, the interpretation of the same image often leads to different results which are not consensual.

We argue that to have more consistent results, it is necessary to develop new methods which take into account all complex types of knowledge used in the image interpretation process. One way to achieve this is to add more semantics in the interpretation process, by using ontologies which describe both image and field knowledge. Indeed, ontologies are a way to facilitate *knowledge sharing and reuse* and can be formalized with standardized languages as OWL. Thus, they can serve to structure the semantic interpretation of images.

This paper is structured as follows. First, we introduce some relevant works in the field of image interpretation. Then, we discuss on how taking into account image and field viewpoints in the formalization process. Section IV is dedicated to explain why the use of geographic standards can be of interest in this context and introduces the theoretical metamodel that we have specified to take into account this kind of information in both viewpoints. Section V presents our first experimental results by focusing on management of spatial relations. Finally, we conclude and give some perspectives to this work.

## II. SOME RELATED WORK IN IMAGE SEMANTICS

Bridging the semantic gap with ontologies is a wide topic which has been studied by many ways.

Ontologies have been used to improve expressivity for both manual and automatic image annotation [1], [2]. Nevertheless, the main purpose of ontologies in image interpretation is to formalize a domain knowledge [3], [4], [5]. Some works mention reasoners and knowledge inference to enrich image information at different steps of image exploitation [6], [7], [8].

Some of other research ways use spatial relationships between objects. For instance we can cite [9] for fuzzy spatial relationships conceptualization in medical images. In [10], RCC-8 based spatial relationships are used for image annotation. At a lower level, [11] uses spatial relationships for identifying the semantic of urban objects by the way of a dedicated algorithm. Finally, [6] uses spatial relationships in image interpretation based on reasoning. This work argues that some academic spatial relationships cannot be distinguished in digital images because of lack of depth information.

 ISBN: 978-1-61208-251-6

## III. IMAGE AND FIELD VIEWPOINTS

In the process of interpretation of satellite images, image and field are two complementary and indivisible viewpoints, that represent the same features according to distinct perspectives. For example, according to respectivly field and image viewpoints, mangrove can be defined by either biotic or abiotic factors of ecosystems such as leaf type, salinity environment or by physical characteristics such as wavelength or texture (cf. Fig. 1).

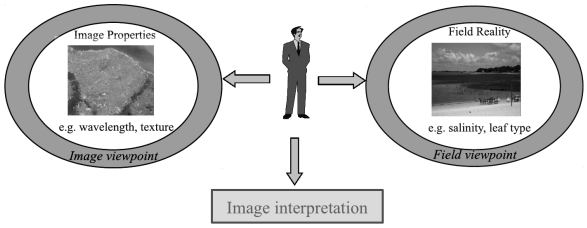



Figure 1. Image and field viewpoints.

However, to be used widely, matching must be established between concepts related to the field and the image points of view (e.g. we must recognize that mangrove defined in the field's point of view apparent to a vegetation segment in the image perspective). To facilitate these gateways, first of all it is necessary to represent knowledge in a common formalism.

Ontologies that [12] defined as a *formal specification of a shared conceptualization*, specify a set of concepts, relations and properties that are relevant for modeling a domain of study [13]. To be used in a process, these ontologies must be formalized in a common language such as the OWL semantic web standard. Furthermore, OWL standard based on description logics, allows reasoning on resources, on the one hand to check the consistency of conceptualized information, and on the other hand to find new knowledge. Formalize knowledge from two points of view by ontologies specified in OWL, will therefore facilitate matching between the different knowledge and allow inferring new knowledge that can be used into a semantic interpretation process.

Moreover, in both points of view, spatio-temporal concepts are commonly used to define features. For example, we can use spatial relations to define features relative position from each other (e.g. Mangrove is located *between* ocean and continent in field, or vegetal segment is *between* water and soil objects in image). Furthermore, geometry can also be used in both viewpoints to define shape or location properties. This consideration also applies with temporal characteristics. Thus, we also must take into account the spatio-temporal dimension in the modeling of knowledge.

As a result, we need to formalize three ontologies: One dedicated to image perspective, another one for field viewpoint and a transversal one to specify spatio-temporal concepts (cf. Figure 2).

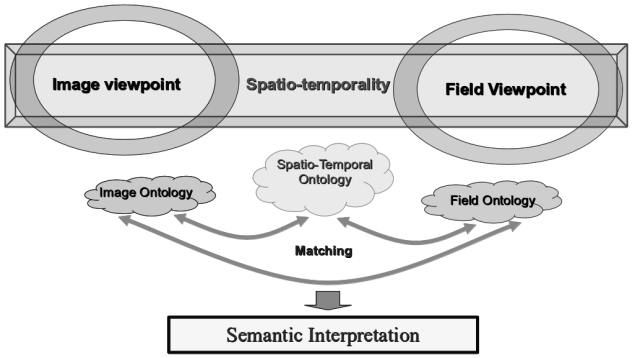



Figure 2. Matching different viewpoints.

## IV. FORMALIZING SPATIO-TEMPORAL CONCEPTS

Because of spatio-temporal concepts exist in both viewpoints, it seems to be useful to formalize them into a framework ontology. Indeed, specifying a framework ontology will give a common basis for describing the different viewpoints, thereby helping the implementation of bridges between the various elements to be described, which, therefore, leads to reduce this which is usually called the *semantic gap* [9]. We also see another advantage which is to unify definitions and to reuse concepts that appear redundantly in different application contexts (e.g. spatial relationships between objects in land cover mapping or Amazonian coastal dynamics).

### A. Using geographic standards to represent spatio-temporal knowledge

GIS community has been very active for many years in terms of modeling spatiotemporal knowledge [14], [15], [16], [17], [18], [19]. A number of these studies have led to the specification of standards and recommendations from the OGC (Open Geospatial Consortium) and ISO (International Organization of Standardization), the two major standard organisms in terms of geographic Information [20], [21], and are now widely used by GIS and spatial DBMS. Thus, it seems interesting to use standards and recognized work to define spatio-temporal concepts.

### B. Spatio-temporal metamodel

A metamodel for describing spatio-temporal knowledge using geographic standards has been defined [22]. An overview of this metamodel is given in Figure 3, where eight components have been identified.

The *Core package* is the central element of the metamodel. It is used to characterize the geographical feature as a whole which have a direct or indirect dependency with the other packages. The *SpatialDimensionPackage* contains information about spatial references of the feature such as shape or location. The *TemporalDimensionPackage* includes concepts, which characterize time such as instant or period.

 ISBN: 978-1-61208-251-6 

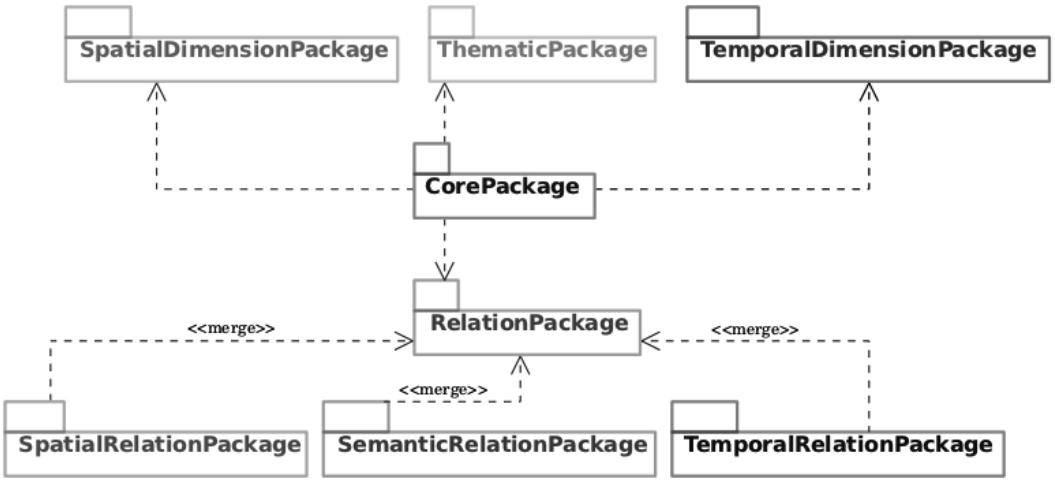



Figure 3. Global view of the metamodel.

The *Thematic package* aim is to describe the other nature of a feature, such as image characteristics or landscape properties. The *Relation package* contains all the required concepts for describing a relationship between features. It is specialized into three sub-packages in order to refine it in terms of *Spatial, Temporal and Semantic relations*. The *TemporalRelation package* includes concepts to define temporal relationships between features, such as ”before” or ”four months ago”. The *SpatialRelation package* includes concepts to define spatial relations between features, such as ”near” or ”50m away”. Finally, the *SemanticRelation package* includes all the others relations that can exist between features such as *part of*, *is a*, *grow*, etc. More details about each package can be found in [22].

For most concrete aspects of the work, our metamodel can be instantiated into a framework ontology which import all ontologies defined by the packages into the metamodel.

### C. Focus on spatial relations

As the experiments given in the next section, focuses on spatial relations, we detail here this package in Figure 4.

Many directions have been taken to define spatial relations, and are currently used in the standards [15], [20], [21], [23], [24]. In our metamodel, we have chosen to use the types defined in [25] to specify three classes of spatial relations. Thus, three types of relation have been defined: *topological, projective and metric*. *Metric relations* are of distances or angles [16]. They can be defined by measurable methods (e.g. the town is located 5km away from the beach), cognitive methods (e.g. forest is near river), or fuzzy methods.*Topological relations* are about connections between objects. These relationships are generally defined by measurable methods (e.g. via the DE9IM matrix [15]), but can also be expressed by terminologically cognitive methods (e.g. next to, touches, within). Three approaches are regularly cited in the literature, namely, the *point set based model of nine intersection* by [15] (*EhRelation*), the *Logic based Model connection calculus regions* of [24] (*RCC8Relation*), and the *Calculus based model* of [23] (*CBMRelation*). We choose to explicitly define these three classes in our metamodel, because they are commonly used by several communities and they can be easily linked to each other [26]. *Projective relations* are described by space projections such as cardinal relationship (e.g. east of, north of) [16], or orientation relations of the objects against each other (e.g. left, down, front) [27]. Finally, we choose to represent reference systems used with the relation by an attribute, whose type is defined in [28] (i.e. intrinsic, extrinsic and deictic).

## V. Experiments on spatial relations

We propose to illustrate relevance of this metamodel integrating some spatial relationships to satellite image interpretation by reasoning. Accordingly, we use spatial relation knowledge by focusing on the RCC8 topological relationships [24].

Our example concerns a calibrated (in reflectance and temperature) Landsat 5-TM image of the surroundings of the city of Santarem (in the Brazilian Amazon) from December 7, 2009. In this case, we attempted to detect segments with different semantics (*cf.* Figure 5).

Image processing is performed using the free software *Orfeo Toolbox* which proposes an implementation of RCC8 spatial relationships. Above all, we ensured the image objects are meaningful. For this purpose, we produced a so-called *good segmentation* [29] based on the preliminary semantic mapping of pixels of [30].

Once the segmentation done, we begin the ontological image description without paying attention to semantics from pixel classification. In the ontological image description approach we detailed in [31], ontologies have two purposes.

 ISBN: 978-1-61208-251-6 

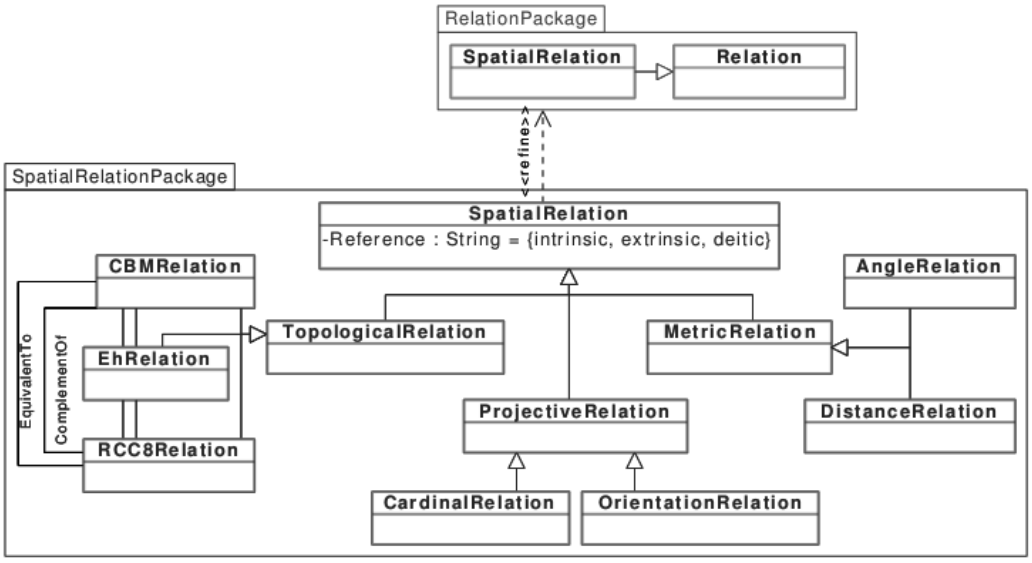



Figure 4. Focus on spatial relations.

Figure 5. Landsat 5-TM calibrated image extract.

(A) (B)

Figure 6. Segments retrieved (in white) by reasoning for two remote sensing semantics: water (A) and built-up and mineral (B).

On the one hand, we use ontologies to automatically describe only low-level image features with a *satellite image reference conceptualization* [6] (framework ontologies). Image description is a set of assertions based on the reference conceptualization, using image concepts, references to remote sensing processing and spatial relationships.

On the other hand, the reference conceptualization is also necessary to express a few expert concepts, which is called *remote sensing contextual knowledge* [6] (domain ontology). We began defining concepts like *vegetal segment*, *water segment* and *mineral segment* using radiometric characteristics (*cf.* Figure 6). These definitions are image-based rather than based on the field point of view.

Concepts like *water segment* and *mineral segment* are defined by intrinsic characteristics which often are spectral or textural indexes. But some other expert concepts need extrinsic characteristics like spatial relationships to be defined. For example, we can define the *beach segment* concept from the mineral one, specifying that a *beach segment* **is a** *mineral segment* **externally connected** to a *water segment*. The RCC8 *externally connected* is an instance member of the spatial relation ontology.

Finally, image segments are automatically classified using a reasoner based on description logics. This allows producing semantics using reasoning based on intrinsic and extrinsic characteristics. The Figure 7 illustrates our approach on the *beach segment* example, taking topological spatial relationships into account.

It is not easy to evaluate the results. We proposed in [32] some confusion matrices to evaluate reasoning on intrinsic characteristics. This is not an absolute solution because the result depends on the whole processing, including segmentation step.

We do not propose quantitative evaluation concerning extrinsic characteristics. However, it is possible to qualitatively compare Figure 5 and Figure 7: the coastline seems to be

 ISBN: 978-1-61208-251-6 

well detected.

Figure 7. Segments of beach retrieved (in white) by reasoning, using spatial relationships.

## VI. Conclusion

We showed in this article how spatio-temporal conceptualization based on reliable standards can be usefully applied to remote sensing.

We presented a proposal of spatio-temporal metamodel and we specifically focused on spatial relations package. We partially create models which conform to the metamodel, using spatial topological relationships for satellite image description.

Finally, we illustrated our work by preliminary example about coastline detection. This concerns the image side and other concrete aspects concerning field point of view are in progress so to include all knowledge areas of remote sensing.

## Acknowledgment

The work presented in this paper was carried out as part of the PO FEDER GUYANE 2007-2013 program under the frame of the CARTAM-SAT project, and of the European Union's Seventh Framework Programme FP7/2007-2013, SPA.2010.1.1-04: ¨Stimulating the development of 490 GMES services in specific area¨, under grant agreement 263435, project BIO_SOS: BIOdiversity Multi-Source Monitoring System: from Space To Species (http://www.biosos.eu), coordinated by CNR-ISSIA, Bari-Italy.

## References

[1] E. Hyvnen, A. Styrman, and S. Saarela, "Ontology-based image retrieval," 2002, pp. 15–27.

[2] M. Srikanth, J. Varner, M. Bowden, and D. Moldovan, "Exploiting ontologies for automatic image annotation," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '05. New York, NY, USA: ACM, 2005, pp. 552–558. [Online]. Available: http://doi.acm.org/10.1145/1076034.1076128[accessed:2012-12-21]

[3] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "An ontology approach to object-based image retrieval," in *In Proc. IEEE Int. Conf. on Image Processing (ICIP03)*, 2003, pp. 511–514.

[4] A. Puissant, N. Durand, D. Sheeren, C. Weber, and P. Ganarski, "Urban ontology for semantic interpretation of multi-source images," in *2nd Workshop Ontologies for urban development: conceptual models for practitioners (Urban Ontologies 2007)*, Castello del Valentino, Turin, Italy, October 2007.

[5] S. Derivaux, G. Forestier, C. Wemmert, and S. Lefevre, "Extraction de détecteurs d'objets urbains à partir d'une ontologie," in *Atelier Extraction de Connaissance à partir d'Images (ECOI), Journées Francophones Extraction et Gestion des Connaissances (EGC 2008)*, Sophia Antipolis, France, Jan 2008, pp. 71–81.

[6] Z. Falomir, E. Jiménez-Ruiz, M. T. Escrig, and L. Museros, "Describing images using qualitative models and description logics," *Spatial Cognition & Computation*, vol. 11, no. 1, pp. 45–74, 2011. [Online]. Available: http://krono.act.uji.es/people/Ernesto/qimage-ontology/paperJSCC-preprintNov2010.pdf/view[accessed:2012-12-21]

[7] B. Neumann and R. Mller, "On scene interpretation with description logics," *Image and Vision Computing*, vol. 26, pp. 82–101, january 2008.

[8] J.-P. Schober, T. Hermes, and O. Herzog, "Content-based image retrieval by ontology-based object recognition," in *in KI-2004 Workshop on Applications of Description Logics*, 2004, pp. 61–67.

[9] C. Hudelot, J. Atif, and I. Bloch. (2008) Fuzzy spatial relation ontology for image interpretation.

[10] N. Maillot and M. Thonnat, "Ontology based complex object recognition," *Image Vision Comput.*, vol. 26, no. 1, pp. 102–113, Jan. 2008. [Online]. Available: http://dx.doi.org/10.1016/j.imavis.2005.07.027[accessed:2012-12-21]

[11] N. Durand, S. Derivaux, G. Forestier, C. Wemmert, P. Ganarski, O. Boussad, and A. Puissant, "Ontology-based object recognition for remote sensing image interpretation," *19th IEEE International Conference on Tools with Artificial Intelligence*, vol. 19, pp. 472 – 479, October 2007.

[12] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," *International Journal of human Computer Studies*, vol. 43, pp. 907–928, 1995.

[13] T. Gruber, "Ontology," in *Encyclopedia of Database Systems*, 2009, pp. 1963–1965.

[14] J. F. Allen, "Maintaining knowledge about temporal intervals," *Commun. ACM*, vol. 26, pp. 832–843, November 1983. [Online]. Available: http://doi.acm.org/10.1145/182.358434

[15] M. Egenhofer, "A formal definition of binary topological relationships," in *3rd International Conference on Foundations of Data Organization and Algorithms*, 1989, pp. 457–472.

[16] A. U. Frank, I. Campari, and U. Formentini, Eds., *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, International Conference GIS Proceedings*, ser. Lecture Notes in Computer Science, vol. 639. Springer, 1992.

 ISBN: 978-1-61208-251-6 

[17] S. Spaccapietra, "Editorial: Spatio-temporal data models and languages," *GeoInformatica*, vol. 5, pp. 5–9, 2001.

[18] L. Spéry, C. Claramunt, and T. Libourel, "A spatio-temporal model for the manipulation of lineage metadata," *GeoInformatica*, vol. 5, no. 1, pp. 51–70, 2001.

[19] C. Parent, S. Spaccapietra, and E. Zimányi, *Conceptual modeling for traditional and spatio-temporal applications - the MADS approach*. Springer, 2006.

[20] ISO/TC211, "Iso geographic information/geomatics technical committees 211," 1994. [Online]. Available: http://www.iso.org/iso/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=54904[accessed:2012-12-21]

[21] OGC/TC, "Opengis abstract specification," 1999. [Online]. Available: http://www.opengeospatial.org/standards/as[accessed:2012-12-21]

[22] C. Pierkot, "A geographic standards based metamodel to formalize spatio-temporal knowledge in remote sensing applications," in *6th International Workshop on Semantic and Conceptual Issues in GIS (SeCoGIS 2012)*, 2012.

[23] E. Clementini and P. D. Felice, "A model for representing topological relationships between complex geometric features in spatial databases," *Information Sciences*, vol. 90, pp. 121–136, 1996.

[24] A. G. Cohn, B. Bennett, J. Gooday, and N. M. Gotts, "Qualitative spatial representation and reasoning with the region connection calculus," vol. 1. Hingham, MA, USA: Kluwer Academic Publishers, October 1997, pp. 275–316.

[25] E. Clementini, "A Conceptual Framework for Modelling Spatial Relations," Phd in Computer Sciences, INSA Lyon, 2009.

[26] M. Perry and J. Herring, "Ogc geosparql, a geographic query language for rdf data," Open Geospatial Consortium, Tech. Rep., 2011, oGC candidate standard.

[27] D. Hernández, *Qualitative Representation of Spatial Knowledge*, ser. Lecture Notes in Computer Science. Springer, 1994, vol. 804.

[28] G. Retz-Schmidt, "Various Views on Spatial Prepositions," *AI Magazine*, vol. 9, no. 2, pp. 95–105, 1988.

[29] G. Castilla and G. J. Hay, *Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications*. Springer, 2008, ch. Image objects and geographic objects, pp. 91–110.

[30] A. Baraldi, V. Puzzolo, P. Blonda, and L. Bruzzone, "Automatic spectral rule-based preliminary mapping of calibrated landsat tm and etm+ images," *IEEE Transactions on geoscience and remote sensing*, vol. 44, no. 9, pp. 2563 – 2586, september 2006.

[31] S. Andrés, "Génie ontologique et raisonneurs : application aux images satellitaires," in *Actes de la conférence Informatique pour les Systèmes d'Information de Décision*, Montpellier, France, May 2012, pp. 575–582.

[32] S. Andrés, D. Arvor, and C. Pierkot, "Towards an ontological approach for classifying remote sensing images," in *SITIS 2012*, 2012.

 ISBN: 978-1-61208-251-6 

# A Methodology for Automatic Analysis and Modeling of Spatial Environmental Data

Mikhail Kanevski
University of Lausanne, Centre for Research on Terrestrial Environment, Geopolis building
1015 Lausanne, Switzerland
Mikhail.Kanevski@unil.ch

*Abstract*—**The research paper deals with a step-by-step methodology for the automatic modeling of geospatial environmental data. The methodology proposed is based on general regression neural networks (GRNN) and probabilistic neural networks (PNN) as modeling tools. GRNN and PNN are nonparametric nonlinear models suitable for the automatic analysis, modeling, and spatial predictions of complex environmental data. The simulated and real data case studies illustrating the methodology are considered and discussed.**

***Keywords-environmental geospatial data; automatic data modelling; machine learning algorithms.***

## I. INTRODUCTION

The problem of automatic environmental data modeling becomes more and more important taking into account the volume of data available from different sources: measurements, automatic monitoring networks, remote sensing, GIS (Geographical Information Systems), etc. These data are widely used to calibrate science-based models (e.g., in meteorology, climate, pollution dispersion), to estimate environmental risks and natural hazards (landslides, avalanches, forest fires etc.), and to estimate renewable resources. Most of environmental data bases contain extremes and outliers and data are highly variable at several spatial scales. Moreover, the environmental phenomena are nonlinear and in many cases should be considered in a high dimensional feature spaces composed of 3d geographical coordinates and additional characteristics derived, for example, from digital elevation models [1][2][3].

A classical approach to analyze environmental geospatial data is based on geostatistical models [4][5]. Most of geostatistical models explicitly take into account the anisotropic spatial correlations analyzed and modeled by an application of variography. In general, geostatistics is a powerful and well established data modeling tool in a low dimensional space.

Recently, an intercomparison of models suitable for automatic two dimensional interpolations in a geographical space was carried out and the results are presented in a report [6]. A wide variety of methods was used – from traditional geostatistical kriging models to advanced neural networks. The general regression neural network produced very good results in terms of testing error and other global statistics usually used to quantify the quality of modeling.

The present study generalizes the ideas proposed in [2][7][8][9] for modeling and predictions of high dimensional complex spatial environmental data. In the following section, a description of a general problem is given and an operational and efficient methodology for automatic geospatial data analysis and modeling is considered step-by-step with a short discussion of real and simulated data case studies. The paper is completed with a brief discussion and some conclusions for the future research.

## II. THE METHODOLOGY AND CASE STUDIES

In general, the problem considered is the following: having environmental data embedded in a high dimensional space, develop a model that 1) explains the phenomena under study without overfitting of data and 2) is good enough (the criteria should be defined) for the generalization and spatial predictions, often called mapping. Moreover, it is necessary to justify the quality of the results obtained by using some general criteria and to quantify the uncertainties of the predictions.

Environmental phenomena are nonlinear and their data modeling should be considered in a high dimensional space (dimension d>3) that is composed of geographical coordinates and some additional variables (features), for example, produced from digital elevation models (e.g., slopes, curvatures, variability at different scales, etc.). Moreover, only some of features could be relevant for the analysis and predictions, and some of them can be just a noise. Therefore, the problem of relevant automatic features selection or features extraction during an automatic analysis can be important.

Taking into account the comments and demands given above, the following generic methodology can be proposed for the spatial environmental data analysis and modeling (from exploratory analysis to spatial predictions and decision-oriented mapping):

1. Preparing of an input/feature space (a collection of independent variables). In principle, the library of input features should be quite general to cover a wide range of possible scenarios.
2. Analysis of monitoring networks and data clustering taking into account the validity domains of raw data and a prediction grid. Monitoring network analysis also helps to understand the representativeness of data and their spatial topology. It improves the decision on data splitting and declustering procedures, and, if necessary, in a monitoring network optimization (redesign) [5][10].

 ISBN: 978-1-61208-251-6 

3. Exploratory spatial analysis of data (ESDA) using visualization (geo-visual analytics), (geo)statistical tools [4][5], and machine learning algorithms (MLA) [8]. At this important stage of the analysis, MLA can help to detect potential patterns in a high dimensional feature space. Data pre-processing and data transformations.
4. Splitting of data into training (development of the model), validation (fitting or calibration of the user-defined hyper-parameters), and testing (assessment of the generalization) subsets. Different criteria can be used: random splitting, spatial declustering, etc.
5. Detection of the available patterns/structures in a high dimensional data. The discrimination between white non-structured noise and spatially structured information. In geostatistics usually the variography is used. In a more general case, MLA can be efficient as well.
6. Training (optimization, calibration) of the models. Modeling of the observed structured information.
7. Iterative application of the feature selection algorithms - either the features are weighted according to their importance or the group of the most relevant features for the prediction is selected [1][2][3][8]. These techniques can also be applied at the steps 1, 5 and 6.
8. Analysis of the training residuals using visualization, (geo)statistical and machine learning tools. Analysis of the residual patterns. The same procedure as in 5 is applied: the residuals should have no spatial structure and should be normally distributed. Moreover, an overfitting of the training data should be avoided. One of the possibilities is to estimate a noise level in data, for example, using an estimate of a nugget in a high dimensional feature space. Then, the variance of a noise can be used as a stopping criterion.
9. Testing of the models. Application of the developed models to testing data subset. Analysis of the testing residuals and their spatial structure. Again, retrained GRNN can be used to perform an exploratory spatial analysis of the testing residuals.
10. Application of the validated and tested models for the spatial predictions (mapping in high dimensional spaces).
11. Quantification of the modeling quality: confidence and prediction intervals.
12. Decision–oriented mapping. At this phase GIS can be widely applied.

Current version of the methodology does not include the recommendations on the monitoring network optimization (MNO). This is a separate but closely related problem. The contemporary reviews on MNO approaches along with space-time environmental data case studies are given in [10].

Recently, the application of general regression neural networks (GRNN) for the regression and the probabilistic neural networks (PNN) for the classification were reconsidered taking into account their properties of patterns detection and adaptivity in the feature selection problems [2][6][7][8]. For example, GRNN is an efficient tool to discriminate noise from structured information. This property can be used both for the original data and for the analysis of the residuals to estimate the quality of modeling. As it was mentioned above, "good" residuals should be white spatially non-structured noise. Basic GRNN model has no hyper-parameters and is easy to train. Therefore, GRNN is attractive model for the automatic data processing.

Anisotropic GRNN can automatically neglect highly noisy features and takes into account only the relevant ones [2][7]. Anisotropic GRNN (when Gaussian kernel is used) means that different kernel bandwidths are used for different features (independent variables). This version of GRNN is sometimes called an adaptive GRNN [2].

It is important to note, that non-parametric statistics is a solid theoretical background both for GRNN and PNN. Therefore, these models can produce also extended results including the characterization of the uncertainties. This is extremely important in a real decision–oriented mapping process when the uncertainties can be even more important than the predictions themselves.

The GRNN/PNN training procedure applied in this research (training = selection of the optimal kernel bandwidths by applying optimization algorithms) is based on a cross-validation error cost function. Either a leave-one-out or a leave-k-out error functions are considered, depending on the number of available training data. In case of too many data a validation data set can be used to train the model. It accelerates the training procedure and reduces the computational time.

The test data set is used only to estimate the generalization properties of the models, i.e., their abilities to predict independent data never seen during the training.

In the present research GRNN was used with an anisotropic Gaussian kernel. In a more general setting a complete Mahalanobis distance can be applied. More theoretical details about the models and their implementation can be found in [8].

The PNN has the same kind of properties and can be used for the classification problems when working with categorical data – discrete classes.

In the present research, the methodology is illustrated using the simulated and real data case studies. Simulated data were produced by adding to the real data several noisy artificial features in order to test the ability of GRNN to neglect the non-relevant information. Artificial additional features were generated using a shuffling procedure, i.e., by randomizing the raw variables. In this case, the original global distributions are preserved but the spatial structures, even if present, are destroyed. The following case studies were considered:

- real data case study. Topo-climatic modeling of the monthly temperature and precipitation in mountainous regions. These are typically three dimensional problems.
- simulated data case study. Three new artificial features were generated either by shuffling of X, Y and Z geographical coordinates or by noise injection

 ISBN: 978-1-61208-251-6 

with different variance. Finally, the problem was considered in a six dimensional space [7].

During the case studies all phases of the methodology were applied. The homogeneity of monitoring network was studied using topological, statistical and fractal measures. Exploratory analysis was carried out using statistical and geostatistical (variography) tools. Measurements were split into training and testing subsets using spatial declustering procedure. Validation subset was not necessary because of the cross-validation (leave-one-out) training technique was applied.

Below the modeling results are discussed briefly. In real data case studies the kernel bandwidths for geographical coordinates reflect spatial 3d anisotropy of the phenomena: for longitude X and latitude Y they are of order 10 km and for altitude Z few hundred meters. The results were compared with a geostatistical model – kriging with external drift [4][8].

In the simulated data case study, noise features after training have very large kernel bandwidths, exceeding the variability of these features. In this case the corresponding part of the Gaussian kernel equals almost to one and these features do not influence the solution. Then, according to the methodology, adaptive GRNN was applied for the exploratory spatial analysis of the training residuals. No spatial structures were detected. At the end, both 3d and 6d (3d+noise) solutions were very similar.

Finally, the models developed were evaluated using testing data subset and good generalization errors were obtained.

More powerful and much more computationally intensive approach is based on a complete analysis of all possible models, i.e. on all possible combinations of features. In a d dimensional input space the number of possible models is $(2^d-1)$. In this case GRNN is applied both as a modeling and as a feature selection tool. Using a cross-validation error, all models can be sorted and the best one with a minimum error can be selected for the predictions. Such approach was also applied for both case studies. Important result is that the best selected models did not include noise features.

## III. DISCUSSION AND CONCLUSIONS

A basic methodology for spatial data processing was proposed. The methodology includes the analysis of input space structure (monitoring networks), comprehensive exploratory analysis of data and the residuals, detection and modeling of structured information using nonlinear nonparametric models. As an efficient and operational tool adaptive GRNN for the regression problems and adaptive PNN for the classification problems were proposed. Training of models was based on cross-validation procedures. The modeling results were evaluated by using independent testing subsets and by analyzing the testing residuals.

One of the important and useful conclusion from the study is that an application of machine learning algorithms at all phases of the data analysis and modeling is strongly recommended [8][9]. In many cases, it helps to reveal complex hidden patterns and structures in data that improves the selection and calibration of models, even if other modeling approaches finally are applied.

The potential extension of the methodology and models can be guided into the following directions: scaling of models with the dimension of space and number of measurements, robustness of the approach, more elaborated assessment of the uncertainties, extension of MLA modeling tools and adaptive kernels [8]. An important future research will be in developing multi-scale multivariate models. For the clustered monitoring networks, kernels can be not only feature-adapted but space-adapted as well.

Only two basic problems of learning from data were considered – classification and regression. The third one and the most difficult – modeling of spatially distributed probability density functions, is still an open question.

Finally, the same kind of methodology can/should be generalized and adapted for the modeling and predictions of spatial-temporal environmental data.

## ACKNOWLEDGMENTS

The author acknowledges useful discussions on the topics presented with Dr. A. Pozdnoukhov, Dr. V. Timonin, Dr. L. Foresti and S. Robert.

The research was partly supported by the Swiss NSF grant #140658.

## REFERENCES

[1] L. Foresti, D. Tuia, M. Kanevski, and A. Pozdnoukhov, "Learning wind fields with multiple kernels", Stochastic Environmental Research and Risk Assessment. Volume 25, Number 1, pp. 51-66, 2011.

[2] S. Robert, L. Foresti, and M. Kanevski, "Spatial prediction of monthly wind speeds in complex terrain with adaptive general regression neural networks", International Journal of Climatology, **32**, DOI: 10.1002/joc.3550, 2012.

[3] A. Pozdnoukhov, G. Matasci, M. Kanevski, and R. Purves, "Spatio-temporal avalanche forecasting with support vector machines", Natural Hazards and Earth System Sciences. Vol. 11, pp. 367-382, 2011.

[4] J-P. Chiles and P. Delfiner, "Geostatistics: Modelling Spatial Uncertainty". John Wiley & Sons, Hoboken, NJ, 2012.

[5] M. Kanevski and M. Maignan, "Analysis and Modelling of Spatial Environmental Data", EPFL Press, Lausanne, 2004.

[6] G. Dubois (Editor), "Automatic Mapping Algorithms for Routine and Emergency Monitoring Data", EU Report No. EUR 21595EN, 152 pp, 2005.

[7] V. Timonin and M. Kanevski, "On Automatic Mapping of Environmental Data Using Adaptive General Regression Neural Network". GIS Research UK annual conference, London, pp. 423-427, 2010.

[8] M. Kanevski, A. Pozdnoukhov, and V. Timonin, " Machine Learning for Spatial Environmental Data. Theory, Applications and Software", EPFL Press, Lausanne, 2009.

[9] M. Kanevski, V. Timonin, and A. Pozdnoukhov, "Automatic Mapping and Classification of Spatial Environmental Data", in Geocomputation, Sustainability and Environmental Planning, B. Murgante, Burusso G., and A. Lapucci, Eds., Springer, pp. 205-223, 2011.

[10] J. Mateu and W. Mueller (Editors), "Spatio-temporal design. Advances in efficient data acquisition", Wiley, 2012.

 ISBN: 978-1-61208-251-6 

# Index Keys Method for Analyses of Urban Public Spaces

## Methodological assumptions

Małgorzata Hanzl
Institute of Architecture and Town Planning
Lodz University of Technology
Lodz, Poland
e-mail: mhanzl@p.lodz.pl

*Abstract*— **The paper proposes the index keys methodology of analysis of urban structures. Few hypotheses are formulated concerning the relation between the crowd of people representing a given culture and the urban settings, which constitute their habitus. The geometrical analyses of public spaces, including the examination of urban closures cross-sections and urban silhouettes, draw upon the writings of Kazimierz Wejchert, widely recognised for his contribution to the theory of urbanism in Poland. Quantitative parameters describing the form of space are introduced, including: central angle, corrugation and regularity. Cultural aspects of given settings, as proved by the results of preliminary descriptive analysis of the former Jewish district in Lodz, require developing methods of description of public spaces.**

***Keywords-urban design; public spaces; urban morphology.***

## I. INTRODUCTION

When looking for the relation between urban structures and the culture of space usage, there are three main issues, which should be considered: (1) physical features, including distribution, shape and size of forms defining the space, (2) the distribution and behaviour of the space's users, which reflect a social order and (3) the flow of human movement, which finds its reflection in the sociometric layout of a given place. Flows are connected with movement/traffic and are related to space, following the definition by Yi Fu Tuan [31]. Concentrations enable contact and communication processes. They are static rather than dynamic, thus place related. Both types are closely interrelated, they inseparably interpenetrate each other. Whenever the human flow stops for a moment concentration occurs, though interrelations require more comfortable conditions to take place, among others: time and spatial arrangement. According to the theory formulated by Lynch [22], flows may be approached as paths and concentrations as nodes. Concentrations tend to a static form, while flows serve mainly as a means of getting to some destination. Taking into consideration mostly their static behaviour, the distribution of people in public spaces reflects social order. Cultural aspects of everyday space usage are reflected first of all by its internal organisation and arrangement.

The paper is organised as follows: after the introduction, which points out at the anthropological understanding of situation and the definition of *habitus*, the main methodological assumptions are presented, with an emphasis on the explanation of index keys concept. Further the case study is discussed, which provides explanation for the development of methodology for detailed description of outdoor spaces. Section IV discusses analytical methodology, introducing several parameters intended for description of urban enclosures. Section V provides conclusions from the paper and presents further steps, which are to be taken in order to verify the presented methodology of analyses.

### A. Anthropological concept of situation

In anthropology situation is defined as a theatre of human activities [24]. Goffman [10, p.18] refers to a situation as to *"the full spatial environment anywhere within which an entering person becomes a member of the gathering that is (or does then become) present"*. Anthropologists developed elaborated theory on ways in which a site is converted into a meaningful 'place', by inscribing human activities into the surroundings. The relationship between people and sites encompasses both: attaching meaning to space and *"recognition and cultural elaboration of perceived properties of environments in mutually constituting ways through narrative and praxis"* [20, p.14]. Schumacher [28] states that the role of architecture is to frame social communication [28, p.414]. Thomas, who introduced the concept of situation in the 1920s, defined it as a *"constellation of the factors determining the behaviour"* [30, p.8] after [28, p.420].

### B. The definition of habitus

The morphological approach [23] refers the above concept to the urban structure introducing the notion of *habitus*. The set of identifiable cues, which may be qualified as culture-specific [25, pp.106-107], and referring to spaces, includes features like: *"quality, size, shape, enclosing elements, paving, barriers, and links, etc."*, requires examination with regard to the distribution of human flows and concentrations and their intensities, and consequently occasions for contacts. Both Gehl [9] and Whyte [34] point at similar rules of use of outside spaces. The territorial distribution and exchange of nonverbal cues serves the communication purpose and usually certain semantics may be attributed to it [10]. The behaviour of a given human group in concentrations reflects its culture. The movement component tends to be more universal and less culture dependent, as Hillier and Hansen [18] claim. The thesis is

 ISBN: 978-1-61208-251-6 

made that the rules, which govern the non-verbal communication component of the human group behaviour are the same ones, which govern the distribution of buildings. They represent the same culture of space usage.

## II. METHODOLOGICAL ASSUMPTIONS

A proposal of the methodology for the analysis of public spaces is presented, based on the writings of Wejchert [32], widely recognised for his contribution to the theory of urbanism. Forgoing morphological descriptions of urban structures based on the analyses of plans, i.e., Conzenian school of urban morphology [33], and on the diachronic characteristics of constructions, i.e., Muratori's tradition [2], do not allow for considerations of this kind. The actual, practice-based approach engages the definition of genius loci, notably in rehabilitation projects. The research tries to develop existing descriptions of urban spaces [8] by applying the anthropological approach as presented by Levi-Strauss [21] to the description of cities and urban structures, following and developing the methodology proposed, among others, by Rapoport [25, pp.106], who defined a comprehensive set of features allowing for drawing characteristics of physical structures, including the culture related ones. Hillier and Hansen [18, p.224] ponder on the method of investigating of encounters as morphic languages, concluding that the aim is to establish, how encounter systems acquire differential properties, which would have different manifestations in space.

### A. *Theory of seeing – index keys concept*

Like in the paintings of Van Gogh, seeing is concentrated around few key points, which define, how a scene is perceived [29]. The analyses should provide observation of processes: flows and forces, and concentrate on their key points. Situations, which are the most important for the definition of cultural character, i.e., the moments of human interactions, particularly attract researchers' attention [15] – they are static rather than dynamic. The methodology of key points, analogue to the anthropological method of making photographs by native observers, who are able to notice the clue activities important for their cultures and often unnoticeable for foreigners, allows for observation of socially meaningful activities, responsible for the formation of a cultural specific environment.

The application of the 'key points' methodology assumes the choice of the most obvious perspectives when observing the environment. In the case of urban spaces, it means choosing these view axes, which provide cross-sections perpendicular to the main axe of a given path. The analyses of the cross-section and of the silhouettes may use, among others, the highly efficient methodology proposed by Gal & Doytsher [7], which allows to replace the Line of Sight (LOS) methodology of extracting silhouettes of groups of buildings. In the proposed methodology the simplest way to extract cross-sections and silhouettes is used, based on orthogonal projection. Although as Gestalt psychology proves in the perception of distant buildings the distances are shortened. Further methodology development should also take into consideration these processes as well as the way in which humans perceive environment, e.g., including perception of meaningful entities [12].

## III. THE CASE STUDY

The case study pertains to the former Jewish district in Lodz, in the 19th century, the district served as a habitat of the multiethnic society, in which Jews constituted a majority [16], [17]. These settings were commonly described as possessing a special 'Jewish' character. This notion is evoked by the form of public spaces, different than in other parts of the city. The study refers mainly to the areas of the Old Town and of the central part of Nowe Bałuty. An attempt has been made to define a certain set of features proper to the area, describing its morphological structure and referring mostly to the shape of public spaces, which is repeated in most Polish towns and neighbourhoods formerly populated by Jews [5]. The examination of the character of public spaces as they are perceived by observers, in the case of scapes, which do not exist in their original form, is based mainly on the analysis of archival photographs. The subject of analysis is first the shape of the public realm itself, in the 2D plan view, the cross-section and street silhouette. Moreover the sequences of views in time and the character of buildings itself should be analysed.

### A. *Shape of public spaces, sequential analysis - General features*

The essential features of the outdoor space, characteristic for the given area refer to the issues of scale and dimensions. The narrowness of streets and presence of numerous slight turns and directional differentiation, providing the notion of concavity, and therefore closing the perspective and assuring perceived and felt closure, are factors favouring direct physical interaction. Gehl [9] points at small dimensions of spaces as favourable for establishing relations. The irregularity of enclosures of streets, their broken line, the apparent lack of precise form, which enlarged the amount of border space, where people stop more willingly than in the centre of an open area, facilitates transactions, presentation of goods, etc. The abundance of such spaces enabled the location of numerous outdoor, commercial furniture. Attracting passers-by, was fostered by the presence of numerous small size elements in the outdoor space, providing sham shelter: stalls, kiosks, stands and presentations encouraging buying. Cullen [3, p.103] describes this phenomenon using the example of a street „cross", the main function of which was to stop pedestrians.

Whyte [34] defines the set of features of outdoor space fostering relations pointing at the location inside of the human flow. Gehl [9, p.150] underlines the role of the corrugation of the edge of space (through the presence of elements of urban equipment and the shape of walls themselves) as important for enhancing communal life.

### B. *The ubiquity of commerce*

The basic character of the area of concern may be defined as the 'ubiquity of commerce'. The space of commerce was not restricted to the main square, it was

 ISBN: 978-1-61208-251-6 

present in the neighbouring streets and passages. Frequent protrusions of buildings, especially of commercial and service use (restaurants, etc) additionally influenced the presence of service in the public sphere, and thus improved the effectiveness of sale. Very rational management of space, lack of space without prescribed use, frequent overlapping and synergy of different uses of the same space completed the above picture. Jews often chose the settlement location in the direct proximity of commercial places. After settling, they usually redeveloped their environment introducing enhancements with regard to the requirements of commerce.

### C. Issues related to proxemics

The proxemics approach, presented by Hall [15] and his successors, examines the relation between spatial patterns of space usage in different cultures and the material environment. The differences between morphological structures representing various cultures are particularly apparent in cities, which like Lodz had become a melting pot of many cultures. Hall [15] identifies direct relationships between interpersonal distances and other characteristics of individuals and communities and the way they shape their own physical environment. Hillier and Hanson [18, p.27] refer to the usage of space and the patterns of behaviour appropriate for different communities and ethnic groups as the determinants of the final shape of urban structures. According to Hillier [19] city is seen as a system of visual distances, strongly influenced by both perception and personal distances.

In social groups, whose members are accustomed to residing in small spaces, social distances are usually smaller than in other groups. Assessment based on the descriptions of the crowd in literature or photos of the Ashkenazi Jewish population, which once used to live in Lodz, corresponds to that characteristic. The limited scale of outdoor spaces, narrowness of the passages and nooks, often even narrowed by the introduction of additional trade facilities, typical for most of former Jewish towns and districts, also fit into this profile. The analyses of crowd with the use of archival photographs, allow to distinguish apparent clusters of people, who grouped also when moving; thus the narrowness of sidewalks.

### D. Perception as a factor influencing the creation of space

Strzemiński [29] pointed at the evolution of visual awareness along with the development of civilisation. Visual awareness was transformed together with the changes of socio-cultural settings. He noticed the result of economic and technical factors as well as the social structure proper to a given group of people, in the defined historical context. The notion of visual awareness, understood as the *"cooperation of seeing and thinking"* emphasises the role of cognitive absorption of perceived visual stimuli. Strzemiński [29] identifies two ways of development of visual awareness. In the rural cultures, it is the observation of the interior of an object, which finds its expression in the studies of nature. The second form was a silhouette vision, which developed from the primitive contour observation in economies based on hunting and breeding animals, that is in tribes accustomed to vast open spaces.

The derivative of the silhouette vision was the perspective of simple parallel projection, and, in the further stage, the development of rhythm, including architectural rhythmisation, as a consequence of inclusion of the afterimage phenomena, natural for the perception processes taking place in vast open spaces. Another form of seeing, which was particularly apparent in communities, whose main occupation was commerce was seeing concentrated on ware attributes, with emphasis on the texture and weight of objects, usually devoid of larger perspective. The preserved iconography, mainly paintings by Jewish artists contemporary to the development of the *'shtetl'* culture, confirms the assumption on their belonging to this group. The shape of urban settings analysed above also confirms the thesis about concentration on the content rather than on external appearance of activities and environment itself.

Adorno [1, p.5] points at the role of artworks as medium reflecting the unconscious aspects of culture: *„Artworks are afterimages of empirical life insofar as they help the latter to what is denied them outside their own sphere and thereby free it from that to which they are condemned by reified external experience."* The same refers to the urban settings, which perceived by a group of users answer their needs, including the aesthetic criteria.

## IV. THE ANALYTICAL METHODOLOGY

### A. Convex - definition

Hillier and Hanson [18] defined a series of rules governing the spatial order of analysed settlements. They noticed that the definition of the basic spatial unit for analyses, which would be distinguishable in the geometrical way, is essential for further considerations. Referring to their theoretical apparatus such basic spatial unit, which may serve for description of public spaces, is a convex. A *"fully convex fat space"* is defined as *"a part of a space, which represents the maximum extension of the point in the second dimension given the first dimension"* [18, p.91]. In Hillier logic of space, the implicit assumption is made that all the cells, representing spaces, are similar units, both in size and in shape. It doesn't describe the actual form of urban closures and the spatial edges are lacking. The critique concerns lack of geometrical description of buildings, which form urban settings, including their size, shape and distribution (e.g., [26]).

Spaces, which are not defined spatially but by the presence of some other edges – like property borders, remain problematic. A more complete picture, which may serve to describe reality in a reliable way, requires introduction of the shape and size parameter(s) and multiplying them by three dimensions. Studies in human perception show a trend to generalise objects to wholes, if only the compounds are located close to each other, have similar attributes, may be described with the same contour line and their meaning, recognised from former experience, remains similar.

 ISBN: 978-1-61208-251-6 

*B. Description of the form of space*

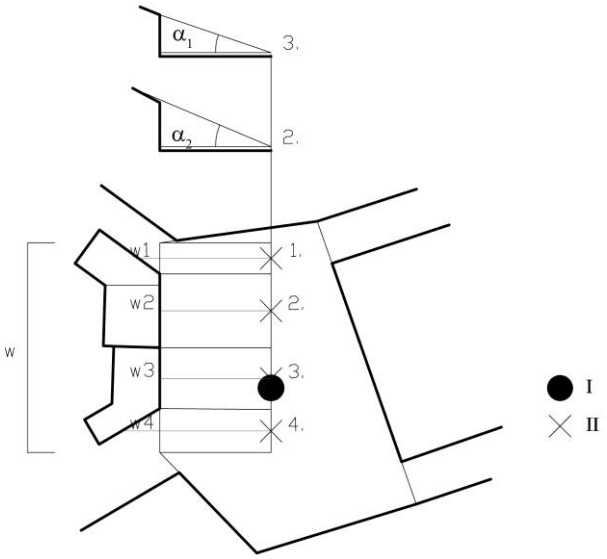



Figure 1. Index points and cross-sections definition. I – geometrical central point, II – index key points, $\alpha_n$ – central angles for each cross-sections, $w_n$ – widths of each unique part of wall

The way in which an observer perceives space in the urban interior depends on the parameters of cross-section. Wejchert [32]. The basic features important for describing convex spaces are cross-section and walls silhouettes. The analysis of a wall silhouette allows for identification of required index points, which may further on serve for creation of cross-sections. Cross-sections may be created for any cue point of any unique physical form of objects surrounding the space change, i.e., the height and the shape of buildings.

Each index point is referred by one cross-section, various cross-sections require association with distinguished index points. The starting point for each cross-section is located on the line, which is parallel to the wall and goes through the geometrical centre of the given convex; see Fig. 2. Cross-sections are by definition perpendicular to the convex wall. In case of buildings or other constructions, which are set back from the convex edge and not perpendicular to it, the middle point of a building/construction is the location of an index point. Similar situation occurs in the case of buildings, which are located behind other buildings but their height exceeds the height of the front building.

The method may also serve for the description of some concavity closures. Yet, as their perception as one spatial unit is more the result of tradition than of their geometrical attributes, these shapes should be defined manually, i.e., divided into two or more basic convexes and then reconsidered as one whole. An example of a concavity space widely recognised as single urban interior is the L-shaped Piazza della Signoria in Florence.

*C. Central angle*

One of the most important parameters describing cross-sections is the central angle. The central angle is an angle between a horizontal plane parallel to the floor at the height of 1,5m (the medium level of sight for humans) and a line going through the highest point of the building defining the closure in a given index point. The point belongs both to the silhouette line and to the cross-section.

Wejchert [32] provides general rules for classification of closures basing on the description of heritage sites which are widely recognised as beautiful for their great proportions. The central angle values in most of the discussed squares range from 25° to 30°, e.g., Piazza Saint Marco in Venice - 28° to 30°, Old Market in Warsaw - 30°. The angle smaller than 10° refers to closures, which are feebly read in space. Either the plan dimensions are too vast or the vertical dimension is not adequate to provide the proper definition of space.

The closures of central angle parameter higher than 60° rarely serve as public piazzas. An important feature for their evaluation are lighting conditions proper for a given climate. The general attitude towards more densely built spaces has changed recently, their values being widely recognised after a break of Modernism. The former pejorative connotation of terms such as “canyon” or “well” [32] lost their previous importance along with common scarcity of defined spaces and dispersion of development. The central angle analysis is made for each of the cross-sections created at each of the index-points of the distinguished walls, and then combined for the walls forming the convexes, using the following formula (1), where $\alpha_1$, $\alpha_2$, $\alpha_3$, $\alpha_n$ are values of central angles of each of the defined cross-sections, n is the number of index points for each wall, $w_n$ is the width of a piece of a wall represented by a given index point and w is the length of the whole wall.

$$\alpha = \alpha_1 \times \frac{w_1}{w} + \alpha_2 \times \frac{w_2}{w} + \alpha_n \times \frac{w_n}{w} = \sum\left(\alpha_n \times \frac{w_n}{w}\right) \quad (1)$$

*D. Corrugation and size*

The urban spaces must be also measured using metric values. Humans, as Gehl asserts in the interview in a documentary film ‘Urbanized’ by Gary Hustwit *“remain a small walking animal”* and require spaces of human scale. The spaces, which are too large seam undefined. Gehl recognises a distance of 100m as a maximum, which allows the observer for proper reception of the environment. The assumed research methodology refers to the width of half of the closure, thus the distance shouldn’t exceed 50m. The actual dimensions of physical spaces reflect also the requirements defined by proxemics. The differences in personal distances influence both the perception of space and its production [13], [15] which means that we may assume that the size of space is perceived and designed differently by people of various cultural background. Continuing this thread the analysis of dimensions of public spaces proves that they remain culture specific.

The definition of space may be either precise or haze. In the first case walls form clearly cut edges, in the second one buildings and other objects are scattered, forming a kind of

 ISBN: 978-1-61208-251-6 

fuzzy boundary. As Wejchert [32] argues, the sight tends towards forms, which are '*strong'* – which means: clearly defined - and towards layouts, which are concise. Parts or whole of the observed constructions may be hidden behind other objects, which occurs both in the vertical as in the horizontal plane. In case of breaks in the structure - i.e., openings in the walls, the closest object closing the perspective visible in the silhouette view is taken into consideration. Similarly a higher building located in the background should be taken into account as, constituting a part of a silhouette, it influences the actual central angle parameter. The index points, where there are no visible constructions are described with central angle value 0.

In a situation, where buildings are set back from the line of frontages, the method allows for the description of an angle in the way similar to other cases. Variations of buildings' set back are another parameter important for the definition of the space character. The line of frontages may be located in the edge of a given convex or set back, the set back may be regular or irregular, any of these attributes influence the perception of the space (Fig. 2).

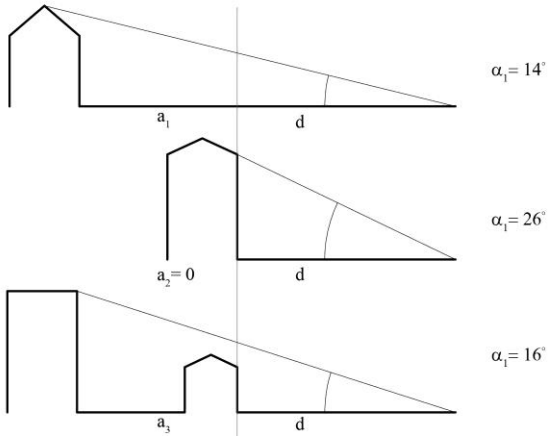



Figure 2. Corrugation of the wall, $a_n$- set back or behind of a part of a wall, d – a distance of the wall from the central point of the cross-section.

Corrugation may be defined using the formula (2), where $\varphi$ symbolises corrugation value of the wall and $\gamma$ - set back of a single part of the wall. The possibilities of comparison of different situations are enabled thanks to the normalisation of set back values as in the formula (3), where a represents the set back in metric units and d – the distance of the wall from the central point of the cross-section. In case of some elements, set behind the lines of frontages shift should be provided as positive numbers.

$$\varphi = \frac{\sum \gamma_n}{n} \quad (2)$$

$$\gamma_n = \frac{a_n}{d} \quad (3)$$

*E. Distribution of index points*

Further analyses include the distribution of index points, which reflects the distribution of buildings – each point belongs to a single building and the points are located in the middle of the facade. Such an analysis allows for easy detection of rhythms, repetitions, symmetries, axial layouts, etc. Distribution of index points may be described as clustered, spaced or scattered. It should be noticed that similar words are applied to the characteristics of the groups of people forming a crowd [6].

When analysing the index points distribution the parameter of regularity may be defined referring to an ideal pattern, which for each case would mean equal distribution of the number of points defined for a given wall (Fig 3). Any shift from the point resulting from an equal division should be measured and normalised by the width of the wall represented by each index point. The sum of all shifts divided by the number of index points describes the value of regularity for each wall. The regularity of the whole closure is described by the average value. The regularity may be described with the use of the formula (4), where τ is the regularity parameter, *r* represents a single shift, *w* – width of a piece of a wall represented by a given index point and *n* is the number of index points for a given wall (5).

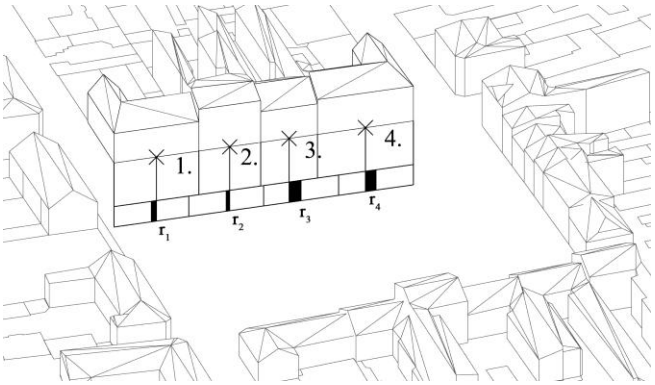



Figure 3. Regularity of the wall – method of description.

$$\tau = \frac{\sum \tau_n}{n} \quad (4)$$

$$\tau_n = \frac{r_n}{w_n} \quad (5)$$

## V. CONCLUSIONS AND FUTURE RESEARCH

The challenge defined by Hall [14, p.55] that in a globalising world man must find out how *"basic cultural systems such as time and space are used to organise behaviour"* starts to influence contemporary urban design thought as numerous studies show [28]. The thread of cultural studies imports viable content to the proposal of

ontology for urban design, which is being developed e.g., by Duarte et al. [4]. The requirement to define the methodology of description of public space character has been recognised. The studies of urban morphology go through a period of intensive revival after a break associated with the activities of modernists [27] and attract the attention of numerous researchers all over the world, as Gauthier and Gilliland [8] describe in their comprehensive résumé.

An extensive set of culture dependent features was defined by Rapoport [25]. The current study provides assumptions to the quantitative description of public spaces basing on the theory by Wejchert [32]. The concept of index points is introduced which enables examination of physical form of urban settings with the use of geometrical description. Basic values are defined, including the parameters of central angle, regularity and corrugation of the walls of the closure. Further development of the current theory is envisaged, including different approaches to the analyses of urban silhouettes and cross-sections, as well as its verification for the description of the assumed case study.

The current research is an ongoing one. The further steps include validation of the proposed methodology in the experimental way and comparison of various urban environments. This may help to understand the diachronic aspects of urban development. The first step assumes exploratory modeling of various urban environments with the use of available software, including: (1) Google SketchUp; (2) CityEngine; (3) Rhino and comparison of results with the two-dimensional analysis explained hitherto.

## REFERENCES


[1] T. W. Adorno, *Aesthetic Theory*. London, New York: Continuum International Publishing Group, 2011.

[2] G. Cataldi, G. L. Maffei, and P. Vaccaro, "Saverio Muratori and the Italian school of planning typology," *Urban Morphology*, vol. 6, no. 1, pp. 3–12, 2002.

[3] G. Cullen, *The Concise Townscape*. Oxford: Elsevier Architectural Press, 2008.

[4] J. P. Duarte, N. Montenegro, and J. N. Beirão, "Public Space Patterns Towards a CIM standard for urban public space," in *Respecting Fragile Places, Proceedings of the 29th Conference on Education in Computer Aided Architectural Design in Europe*, 2011, pp. 79–85.

[5] R. Dylewski, "Lubuski Orient," *Scriptores*, vol. 2, 2003.

[6] N. Fridman and G. A. Kaminka, "Towards a Cognitive Model of Crowd Behavior Based on Social Comparison Theory," 2007.

[7] O. Gal and Y. Doytsher, "Fast and Accurate Visibility Computation in a 3D Urban Environment," in *GEOProcessing 2012: The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services*, 2012, no. c, pp. 105–110.

[8] P. Gauthier and J. Gilliland, "Mapping urban morphology : a classification scheme for interpreting contributions to the study of urban form," *Urban Morphology*, vol. 10, no. 1, pp. 41–50, 2006.

[9] J. Gehl, *Life Between Buildings, Using Public Space*. Kopenhagen: The Danish Architectural Press, 2001.

[10] E. Goffman, *The Presentation of Self in Everyday Life*, Monograph. Edinburgh: University of Edinburgh, Social Science Research Centre, 1959.

[11] E. Goffman, Behavior in Public Places, Notes on the Social Organization of Gatherings. New York: The Free Press, 1966.

[12] S. Guberman, V. V. Maximov, and A. Pashintsev, "Gestalt and Image Understanding," *Gestalt Theory*, vol. 34, no. 2, pp. 143–166, 2012.

[13] E. T. Hall, *Hidden Dimension*. Garden City, NY: Doubleday, 1966.

[14] E. T. Hall, *Beyond Culture*. New York: Anchor Books, 1989.

[15] E. T. Hall, "Proxemics," in *The Anthropology of Space and Place, Locating Culture*, S. M. Low and D. L. Lawrence-Zuniga, Eds. Oxford: Blackwell Publishing, 2009, pp. 51–73.

[16] M. Hanzl, "Urban Structure as a Repository of Social Content – the Case Study of the Lodz 'Jewish District'," in *REAL CORP Proceedings/Tagungsband*, 2012.

[17] M. Hanzl, "W poszukiwaniu śladów dawnej dzielnicy żydowskiej w Łodzi," *Kultura Enter*, vol. 38, no. 9, vol. 40, no. 11, 2011.

[18] B. Hillier and J. Hanson, *The Social Logic of Space*, Cambridge: Cambridge University Press, 1984

[19] B. Hillier, "The genetic code for cities – is it simpler than we thought?," in *Complexity Theories of Cities have come of Age, TU Delft*, 2009, no. September.

[20] D. L. Lawrence and S. M. Low, "Locating Culture," in *The antropology of Space and Place, Locating Culture*, S. M. Low and D. L. Lawrence-Zuniga, Eds. Oxford: Blackwell Publishing, 2009, pp. 1–47.

[21] C. Levi-Strauss, *Antropologia strukturalna*. Warsaw: Wydawnictwo Aletheia, 2009.

[22] K. Lynch, *The Image of the City*. Cambridge: The MIT Press, 1960.

[23] P. Panerai, J. Ch. Depaule, and M. Demorgon, Analyse urbaine, Marseille: Édition Parenthèses, 2009

[24] R. S. Perinbanayagam, "The Definition of the Situation: An Analysis of the Ethnomethodological and Dramaturgical View," *The Sociological Quarterly*, vol. 15, pp. 521–541, 1974.

[25] A. Rapoport, The Meaning of the Built Environment. A Nonverbal Communication Approach. Tuscon: The University of Arizona Press, 1990.

[26] C. Ratti, "Space syntax: some inconsistencies," *Environment and Planning B: Planning and Design*, vol. 31, no. 4, pp. 487–499, 2004.

[27] J. Rykwert, The Idea of a Town, The Anthropology of Urban Form in Rome, Italy and the Ancient World. Cambridge, London: The MIT Press, 1989.

[28] P. Schumacher, *The Autopoiesis of Architecture A New Framework for Architecture*, vol. 1. Great Britain: A John Wiley and Sons, Ltd, Publication, 2011.

[29] W. Strzemiński, *Teoria Widzenia*. Kraków: Wydawnictwo Literackie, 1974.

[30] W. I. Thomas, Primitive Behavior: An Introducton to the Social Sciences. New York: McGraw-Hill, 1937.

[31] Y.-F. Tuan, *Space and Place, The Perspective of Experience*. Minneapolis, London: University of Minnesota Press, 1977.

[32] K. Wejchert, *Elementy kompozycji urbanistycznej*. Warszawa: Wydawnictwo Arkady, 1984.

[33] J. W. R. Whitehead and P. J. Larkham, Eds., *Urban Landscapes, International Perspectives*. London: Routledge, 2000.

[34] W. Whyte, *The Social Life of Small Urban Places*. New York: Project for Public Spaces, 2009.


 ISBN: 978-1-61208-251-6 

# The spatio-temporal semantics from a perdurantism perspective

Benjamin Harbelot*, Helbert Arenas†, Christophe Cruz‡
Laboratoire Le2i, UMR-6302 CNRS,Departement Informatique
University of Burgundy
Dijon, France
*benjamin.harbelot@checksem.fr, †helbert.arenas@checksem.fr, ‡christophe.cruz@u-bourgogne.fr

***Abstract*—In this paper we present the "continuum model". Our work follows a "perdurantism" approach and is designed to handle dynamic phenomena extending the 4D-fluent with the use of semantic web technologies. In our approach we represent dynamic entities as constituted by time slices each with semantic, geometric and temporal components. Our model is able to link the diverse representations of an entity and allows the inference of qualitative information from quantitative one. The inference results are later added to the ontology in order to enhance the knowledge base. The model has been implemented using OWL and SWRL. Our preliminary results are promising and we plan to further develop the model in the near future to increase the suitable data sources.**

***Keywords*-spatio-temporal; semantics; GIS; perdurantism.**

## I. INTRODUCTION

For the design of a spatio-temporal knowledge system, it is necessary to consider the three components of an entity representation: 1) Spatial: consisting in the geometry, 2) Temporal: which defines the interval of existence of the geometries and finally 3) Semantic: which defines a meaning for the entity beyond the purely geographic one [1]. Currently available GIS tools lack the capacity to perform inference or reasoning on information from spatio-temporal dynamic phenomena. An alternative to classic GIS tools are Semantic Web technologies, tools specifically designed to perform reasoning and inference. In this research we use Semantic Web technologies to develop the "continuum model", an ontology that allows us to represent diverse dynamic entities and analyse their relationships along time. Traditionally ontologies are static in the sense that the information represented in them does not change in time or space. In this paper we introduce the continuum model, an ontology that extends the 4D-fluent providing it with the required capabilities to keep track of spatial and semantic evolution of entities along time.

In Section II we discuss related work in the field of spatio-temporal knowledge representation. In Section III we introduce the continuum model, we present the model specification using description logics, in Section IV we describe how the model operates using an example and later we indicate our conclusions and future work.

## II. RELATED WORK

The development of a spatial-temporal knowledge system involves two aspects, first the representation of the knowledge and second, the necessary mechanisms to perform analysis and querying.

### A. Representing temporal data

The two main philosophical theories concerning the representation of object persistence over time are: *endurantism* and *perdurantism*. The first one, *endurantism*, considers objects as three dimensional entities that exist wholly at any given point of their life. On the other hand, *perdurantism*, also known as the four dimensional view, considers that entities have temporal parts, "time slices" [2]. From a perdurantism point of view the temporal dimension of an entity is composed by all its time slices. It therefore represents the different properties of an entity over time as *fluent*. A *fluent* is a property valid only during certain intervals or moments in time. From a designer point of view, the *perdurantism* approach offers advantages over the *endurantism* allowing richer representations of real world phenomena [3].

The implementation of a *perdurantism* approach within an ontology, requires the conversion of static properties into dynamic ones. The two primary Semantic Web languages are OWL and RDF, unfortunately both of them provide limited support for temporal dynamics [4]. The OWL-Time ontology describes the temporal content of web pages and temporal properties of web services. Moreover, this ontology provides good support for expressing topological relationships between times or time intervals, as well as times or dates [5]. However OWL allows only binary relations between individuals. In order to overcome this limitation several methodologies have been proposed for the representation of dynamic objects and their properties. Among the most well known are the temporal description logic, temporal RDF, versioning, reification, N-ary relationships and the 4D-fluent approach.

Temporal RDF [6] proposes an extension of the standard RDF for naming properties with the corresponding time interval. This allows an explicit management of time in RDF. However Temporal RDF uses only RDF triples, therefore it does not have all the expressiveness of OWL for instance,

 ISBN: 978-1-61208-251-6 

it is not possible to employ qualitative relations. Reification is a technique used to represent n-ary relations, extending languages such as OWL that allow only binary relations [7]. In [4], the authors developed a lightweight model using Reification. The model is designed to be deployed on top of existing OWL ontologies extending their temporal capabilities. The model also implements a set of SWRL (Semantic Web Rule Language) operators to query the ontology. Reification allows the use of a triple as object or subject of a property. But this method has also its limitations, for instance the transformation from a static property into a dynamic one increases substantially the complexity of the ontology, reducing the querying and inference capabilities. Additionally reification is prone to redundant objects which reduces its effectiveness. Versioning is described as the ability to handle changes in ontologies by creating and managing multiple variants of them [8]. However, the major drawback of Versioning, is the redundancy generated by the slightest change of an attribute. In addition, any information requests must be performed on multiple versions of the ontology affecting its performance.

The 4D-fluent approach is based on the *perdurantism* philosophical approach.It considers that the existence of an entity can be expressed with multiple representations, each corresponding to a defined time interval. In the literature 4D-fluent is the most well known method to handle dynamic properties in an ontology. It has a simple structure allowing to easily transform a static ontology into a dynamic one although it has some limitations [9]. The 4D fluent approach allows the recording of frequent time slices but it can not handle explicit semantics. This fact causes two problems: 1) It is difficult to maintain a close relationship between geometry and semantics; and 2) It increases the complexity for querying the temporal dynamics and understanding the modelled knowledge. Furthermore, this approach does not define qualitative relations to describe the type of change that has occurred or to describe the temporal relationships between objects. We cannot then know which entities have undergone a change and what entities might be the result of that change. Regardless of its limitations the 4D-fluent approach offers a solid starting point for the representation of temporal information in OWL. A work based on 4D-fluent is SOWL, which extends the ontology OWL-time making it able to handle qualitative relations between intervals, such as “before” or “after” even with intervals with vague ending points [10].

### B. Querying the ontology

Traditionally SPARQL has been the most common language to query an ontology. SPARQL is a W3C recommendation that operates at the level of RDF graphs. However, the queries become relatively complex in a space-temporal system. An extension of this language, st-SPARQL [11], defines new functions that allow it to handle geometries but not temporal data. St-SPARQL is based on an extension of RDF called st-RDF that integrates contact geometries and incorporates time in RDF. St-SPARQL and SPARQL are both based on RDF graphs, therefore it is impossible to draw any inference with them.

In [12] the authors introduce a model in which spatial-temporal information contained in a database and a spatial-temporal inference system work together. However, no information is given on the Semantic Web technologies, only the Java language is quoted as a component of the inference engine, therefore the universality and effectiveness of the inference system can be questioned. Another work is [13] in which the authors propose a reasoning system that combines the topological calculus capabilities of a GIS and the inference capabilities of the semantic web field. However the notion of time is not incorporated in this model.

The capability of switching from quantitative to qualitative data is only possible with a reasoning system. In the case of SOWL this is possible thanks to the implementation of SWRL built-ins. In SOWL the built-ins allow the system to infer topological, directional and metric relations between entities. Qualitative information can be inferred from quantitative one and can be used as an alternative in the case of missing quantitative data. In order to query the ontology the developers of SOWL implemented a language similar in syntax to SQL. This language performs simple spatial-temporal querying for both static and dynamic data [10].

Our literature review suggest us that the most suitable approach to develop a spatial-temporal knowledge system should follow a 4D-fluent approach using SWRL built-ins to perform complex queries and reasoning. In the next section we will describe how we implemented this approach in the continuum model.

## III. The Continuum Model

The 4D-fluent approach does not allow an entity to change its nature, only allows the change of the value of some of its properties. However the semantics associated with a geometry may change. For example a land parcel may change from being forest into being urban. In this example the geometry has not changed, however there is a semantic change (See figure 1A). It is equally possible that the semantics might not change while the geometry evolves. For instance, a given urban land parcel might expand by purchasing neighbouring parcels (see figure 1B).

In order to represent a dynamic entity in the continuum model we create a set of object time slices, each constituted by three components as depicted in figure 2A: 1)Semantic: To describe the knowledge associated with the entity. 2) Spatial: It is the graphical representation. 3) Temporal: It represents the interval or time instants that describe the temporal existence. The goal of the continuum model is to follow the evolution of entities though time. To achieve this goal the model records the changes that entities might

 ISBN: 978-1-61208-251-6 

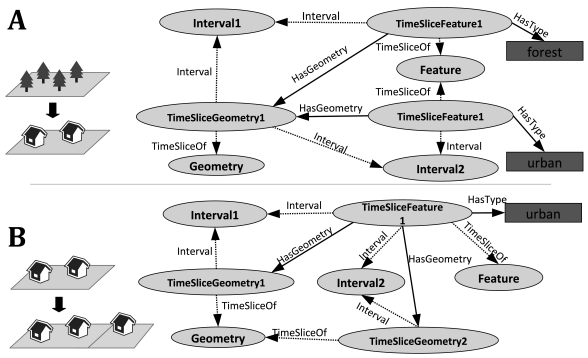



Figure 1. Examples of the evolution A) Two different semantic objects for the same geometry. B) Two related geometries for the same semantic object.

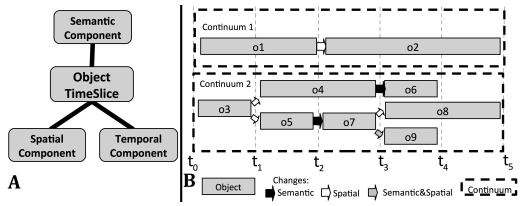



Figure 2. A) The three components of an entity within the continuum model. B) Using the continuum model to represent the evolution of an entity.

go through in their semantic or spatial components along time. For this purpose the model creates a new component representation every time a change occurs (spatial or semantic). The resulting child object retains all the remaining characteristics from the original parent object. Each change adds to the genealogy of the spatio-temporal objects. The parent-child relation is recorded in the system, allowing the analysis and querying of the information. The model enforces a coherency between the time intervals of objects contained in the system.

Figure 2B depicts an example of objects genealogy. In this example objects "o4" and "o5" are children of object "o3", and are the result of an spatial change in the parent object. The system enforces temporal coherency, children objects can not occur before the parent interval. It is possible to characterize the evolution of each object in the model according to the conceptual hierarchy depicted in figure 3.

The continuum groups related objects, which have a valid

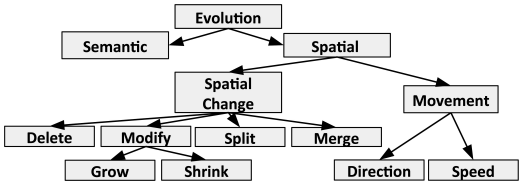



Figure 3. Qualification of transition in the spatial graph.

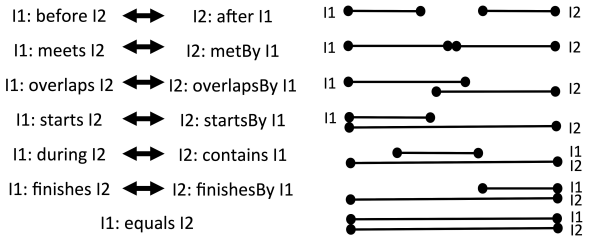



Figure 4. Allen temporal relations.

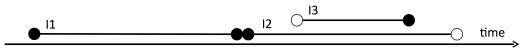



Figure 5. Using Allen temporal relations to infer new knowledge.

time interval of existence.The model links individual objects to their context. For instance an object can belong to more than one continuum, therefore continuums can intersect. Our system allows the definition of qualitative relations between spatial-temporal objects, even when this object belong to different continuums. Figure 2B depicts the evolution of an entity and how the continuum concept is used to study it.

In our model we have implemented qualitative temporal relations based on binary and mutually exclusive relations as proposed by Allen [14] (see figure 4). The addition of Allen relations increase the expressive power of the system by adding qualitative information in addition to the quantitative one. By using defined Allen relations between intervals we can obtain qualitative information even from intervals with vague endpoints in a similar fashion to [9]. For example, figure 5 depicts intervals "I1", "I2" and "I3". While we know the start and ending points of "I1", we do not know the ending point of "I2", and we do not know the starting point of "I3'. However we know that "I1" meets "I2" and that "I2" contains "I3". Then we can infer that because "I2" contains "I3",then "I3" must be after "I1", even if the information about start and ending points is incomplete. Lack of knowledge caused by semi closed intervals is largely filled by the integration of Allen relations to the model.

In GIS, objects or regions are represented by points, lines, polygons or other more complex figures based on these geometries. All these geometries are defined using the coordinates of points which are quantitative information. There are mainly three types of relationships between geometries: directional, metric, and topological relationships. The topological analysis between two objects is done using the models: Dimensionally Extended Nine-Intersection Model (DE-9IM) or RCC8 [15]. In both cases, we obtain an equivalent set of topological relationships for specific regions. To calculate the spatial relationships between two geometries the DE-9IM model takes into account the inside, the outside, and the contour of the geometries leading to the analysis of nine intersections as described in [15] .

There are eight possible spatial relationships of the result-

 ISBN: 978-1-61208-251-6 

Table I
TOPOLOGICAL PREDICATES AND THEIR CORRESPONDING MEANINGS.

| Topological | Predicate Meaning |
|---|---|
| Equals | The Geometries are topologically equal. |
| Disjoint | The Geometries have no point in common. |
| Intersects | The Geometries have at least one point in common (the inverse of Disjoint). |
| Touches | The Geometries have at least one boundary point in common, but no interior points. |
| Crosses | The Geometries share some but not all interior points, and the dimension of the intersection is less than that of at least one of the Geometries. |
| Overlaps | The Geometries share some but not all points in common, and the intersection has the same dimension as the Geometries themselves. |
| Within | Geometry A lies in the interior of Geometry B |
| Contains | Geometry B lies in the interior of Geometry A (the inverse of Within) |

ing analysis-9IM (see table I).

The relationships based on quantitative information can be translated later into qualitative data [16], in a similar fashion as we have described for the temporal aspect. By analysing a set of moments and time intervals it is possible to deduce qualitative topological relationships between objects.

In this section we use a Tarski-style specification to describe the model main components.

To represent time intervals we follow the semantics suggested by Artale and Franconi (1998). We can think of the temporal domain as a linear structure $\mathcal{T}$ composed by a set of temporal points $\mathcal{P}$. The components of $\mathcal{P}$ follow a strict order $<$, which forces all points between two temporal points $t_1$ and $t_2$ to be ordered. By selecting a pair $[t_1, t_2]$ we can limit a closed interval of ordered points. The set of interval structures in $\mathcal{T}$ is represented by $\mathcal{T}^\star_<$ [17].

Temporal Points:

$\mathcal{P} \qquad \mathcal{P}^\mathcal{I} \subseteq \Delta^\mathcal{I}$

Time Intervals:

$\mathcal{T}^\star_< \qquad [to, tf] \doteq \{x \in \mathcal{P} | to \leq x \leq tf, to \neq tf\} in \mathcal{T}$

To define the relations identified by Allen [14] (See figure 4) we first define two intervals $i1$ and $i2$: $\mathcal{T}^\star_<(i1)$ , $\mathcal{T}^\star_<(i2)$, being $i_{to}$ the starting point and $i_{tf}$ the ending point of the intervals.

$Before(i1, i2) \quad (i1_{tf} < i2_{to})$
$Meets(i1, i2) \quad (i1_{tf} = i2_{to})$
$Overlaps(i1, i2) \quad (i1_{tf} > i2_{to}) \wedge (i1_{tf} < i2_{tf})$
$Starts(i1, i2) \quad (i1_{to} = i2_{to}) \wedge (i1_{tf} < i2_{tf})$
$During(i1, i2) \quad (i1_{to} > i2_{to}) \wedge (i1_{tf} < i2_{tf})$
$Finishes(i1, i2) \quad (i1_{to} > i2_{to}) \wedge (i1_{tf} = i2_{tf})$
$Equasl(i1, i2) \quad (i1_{to} = i2_{to}) \wedge (i1_{tf} = i2_{tf})$

The Spatial representation of an object ($\mathcal{SR}$) is composed by a spatial reference system ($SRS$) and a geometry ($\mathcal{G}$) (A more complex definition is possible, however for the sake of simplicity we will refer only to the essential components of a geographic feature definition).

Spatial Reference System: As defined by the European Petroleum Standards Group (EPSG) [18]

$SRS^\mathcal{I} \subseteq \Delta^\mathcal{I}$

Geometries: A set of coordinates that define points, lines, curves, surfaces and polygons.

$\mathcal{G}^\mathcal{I} \subseteq \Delta^\mathcal{I}$

$\forall HasSRS.SRS \equiv \{x \in \Delta^\mathcal{I} | \forall s.(x, s) \in HasSRS^\mathcal{I} \rightarrow s \in SRS^\mathcal{I}\}$

$\forall HasGeom.\mathcal{G} \equiv \{x \in \Delta^\mathcal{I} | \forall g.(x, g) \in HasGeo^\mathcal{I} \rightarrow g \in \mathcal{G}^\mathcal{I}\}$

Then the spatial representation can be defined as:

$\mathcal{SR} \equiv \forall HasSRS.SRS \sqcap \forall HasGeom.\mathcal{G}$

The spatial relations between geometries are defined by the Extended Nine-Intersection model (DE-9IM) [15].

The semantic component of the objects is represented by $\mathcal{S}$. It describes the nature of the entities and can be composed by one or more alphanumeric properties.

Each object time slice ($\mathcal{O}$) in the continuum model has three components: 1) a time interval ($\mathcal{T}^\star_<$), 2) a spatial representation ($\mathcal{SR}$) and 3)a semantic component ($\mathcal{S}$).

$\mathcal{O} \equiv \quad \forall HasSR.\mathcal{SR} \sqcap \forall HasInterval.\mathcal{T}^\star_< \sqcap \forall HasSemDef.\mathcal{S}$

In the continuum model a change on the spatial representation or on the semantic component generates a new object which has a *child - parent* relationship with the original object, additionally we know that the time interval of the parent object *meets* the time interval of the child object (see figure 4). The parent child relationship between object $o1$ and $o2$ is defined by the relationships between their spatial representations ($o1_{sr}$ and $o2_{sr}$), their semantic definitions ($o1_s$ and $o2_s$) and their time intervals ($o1_i$ and $o2_i$)

$\forall HasChild.\mathcal{O} \quad \{o_1 \in \mathcal{O}^\mathcal{I} | \forall o2.(o1, o2) \in HasChild^\mathcal{I} \rightarrow o2 \in \mathcal{O}^\mathcal{I} \wedge \exists((o1_{sr} \neq o2_{sr}) \vee (o1_s \neq o2_s)) \wedge (meets(o1_i, o2_i))\}$

where: $\{o1, o2\} \in \mathcal{O}$ , $\{o1_{sr}, o2_{sr}\} \in \mathcal{SR}$ and $\{o1_s, o2_s\} \in \mathcal{S}$

The spatial transitions in the model are a subset of the $HasChild$ relationship: $SpatialEvolution \sqsubseteq HasChild$ We have implemented the following spatial transitions: (see table I for a definiton of topological relations)

$Merge(input, output)$
$input = \{a_1, a_2..a_n\} | \forall x \in input \rightarrow \mathcal{SR}(x)$
$\mathcal{SR}(output) \wedge output = (a_1 \cup a_2 \cup \ldots a_n)$

$Split(input, output)$
$\mathcal{SR}(input) \wedge output = \{a_1, a_2..a_n\} | \forall x \in output \rightarrow SR(x)$

$Equals(input, output)$

$Delete(input, output)$
$\mathcal{SR}(input) \wedge Equals(output, \emptyset)$

$Grow(input, output)$
$\{input, output\} \in \mathcal{SR} \wedge Within(input, output)$

$Shrink(input, output)$
$\{input, output\} \in \mathcal{SR} \wedge Contains(input, output)$

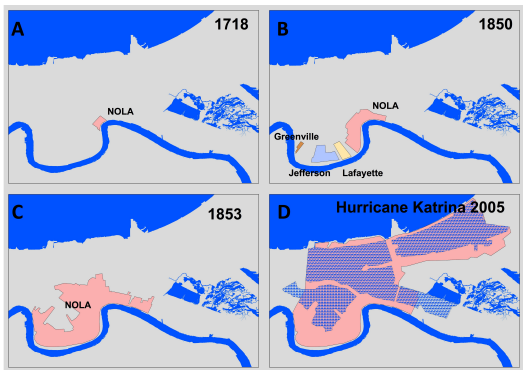



Figure 6. City of New Orleans along time.

## IV. Example Continuum

The continuum model is flexible enough to be adapted in multiple fields. For this example we will use it to study the urban evolution of the city of New Orleans, Louisiana. Figures 6 and 7 represent the urban evolution of the entity *"city of New Orleans"*. Each one of its multiple representations along its history is a *time slice*. Figure 6 depicts the urban evolution of the city, we can see the historic French Quarter, founded in 1718 (Figure 6A), and how the city grew until around 1853 (Figure 6C) when it went trough a conurbation process with the cities of Greenville, Jefferson and Lafayette (Figure 6B). The city continued its growth and by 1949 it reached its approximately modern size. In August of 2005 Hurricane Katrina landed near the city causing a major flood, also depicted in figure 6D. First we define the class Human Settlement ($\mathcal{HS}$) as a subclass of the objects ($\mathcal{O}$), $\mathcal{HS} \sqsubseteq \mathcal{O}$ therefore it has all three components, spatial ($\mathcal{SR}$), temporal ($\mathcal{T}$) and semantic ($\mathcal{S}$).

The conurbation process involves two cities merging. Using the model we can represent the process as:

$\{a1, b\} \in \mathcal{HS} | Contains(a1_{sr}, b_{sr}) \wedge Meets(a1_i, b_i)$
$\rightarrow ConUrbation(a1, b)$

Figure 9 depicts the form how the model will be used in the New Orleans example. The spatial representation of the time slice *Nola2* contains the spatial representation of *Jefferson*, therefore there is a conurbation process by the year 1853.

Figure 6D depicts the area flooded by Hurricane Katrina in 2005.We can create a new class *risk areas* as $\mathcal{RA}$ ($\mathcal{RA} \sqsubseteq \mathcal{O}$), representing the flooded area. Then we can identify the process *growth in risk area* as:

$\{a, b\} \in \mathcal{HS} \wedge r \in \mathcal{RA} | Grow(a, b) \wedge$
$(Overlaps(a_{sr}, r_{sr}) = \emptyset) \wedge (Overlaps(b_{sr}, r_{sr}) \neq \emptyset)$
$\rightarrow GrowthInRiskArea(a, b)$

## V. Conclusion

Figures 8 depicts the representation of the urban growth using the classic 4D-fluent approach, while figure 9 depicts the continuum model. In the later, following the approach by Welty and Fikes, classes *TimeSlice* and *TimeInterval*

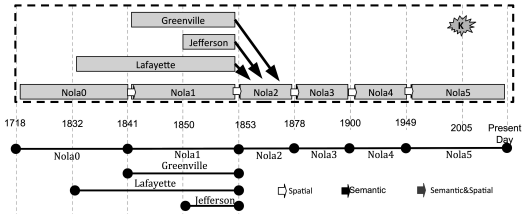



Figure 7. Time frame or urban evolution

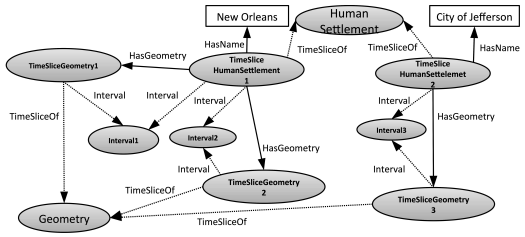



Figure 8. Representation using the 4D fluent

with properties *TimeSliceOf* and *HasInterval* are introduced to allow the ontology to handle temporal entities. Class *TimeSlice* is the domain class for entities representing temporal parts and class *Interval* is the domain class of intervals. A time interval holds the temporal information of a timeslice. Property *TimeSliceOf* connects an instance of class *TimeSlice* with an entity, and property *HasInterval* connects an instance of class *TimeSlice* with an instance of class *Interval*. Our model enhances the understanding of the data represented in the ontology. First, we removed the notion of *TimeSlice* which does not refer to any object in the real world. *TimeSlice* are replaced by instance from explicit Class providing an explicit semantic. Moreover, the 4D-fluent approach is enhanced by adding several types of qualitative relations.Temporal Allen relations and spatial relationships resulting from analysis-9IM. Understanding data semantics is at the core of our work providing an easier way to manage data and reduces queries complexity. When using, reasoning capabilities specific to the web semantic, the system may enrich itself the knowledge store in the ontology. Our model offers explicit semantic and flexibility for semantics interoperability between information systems

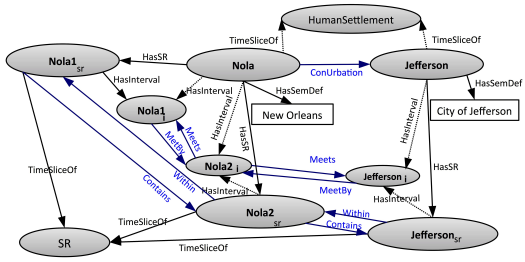



Figure 9. Representation using the continuum model

 ISBN: 978-1-61208-251-6 

and data sharing.

The rules are executed via a graphical interface using the Jena API to connect to the ontology and JDBC to access the database. The application automatically detects the presence of spatial built-ins in SWRL rule and automatically starts the calculation in the database. However, we note two limitations to this model: 1) the treatment of a query containing a spatial built-ins can be very long depending on the number of geometry involved in spatial analysis, 2) the execution of SWRL rule containing spatial built-ins is currently dependent on our application and cannot be executed, for example, from traditional plugin SQWRL Tab Query of the Protégé tool.

A spatial built-ins uses quantitative data to launch a spatial analysis which establishes qualitative relations between the geometries involved in the calculation. Currently, our system can automatically rewrite SWRL rules containing spatial built-ins. On one hand, this allows not repeating the calculations that have already been performed. On the other hand, it also provides a SWRL rule no longer containing spatial built-ins but rather a qualitative relationship expressed through a property defined in the ontology. For example, this rules asking for people within a restaurant: $feat : restaurant(?x) \wedge feat : people(?y) \wedge spatialswrlb : Within(?x, ?y) \rightarrow sqwrl : select(?x)$

will be rewritten as:

$feat : restaurant(?x) \wedge feat : people(?y) \wedge sa : HasWithin(?x, ?y) \rightarrow sqwrl : select(?x)$

However, the addition of new objects in the ontology as well as in the spatial database can make the result of the rewritten query incomplete. It should be necessary to restart a calculation with a spatial built-in to update the qualitative relationships between geometries.

## REFERENCES


[1] M. Yuan, "Use of a three-domain representation to enhance GIS support for complex spatial-temporal queries," *Transactions in GIS*, vol. 3, pp. 137–159, March 1999.

[2] C. Welty and R. Fikes, "A reusable ontology for fluents in OWL," in *Proceedings of 2006 conference on Formal Ontology in Information Systems (FOIS 2006)*, 2006, pp. 226–236.

[3] M. Al-Debei, M. Mourhaf Al Asswad, S. de Cesar, and M. Lycett, "Conceptual modelling and the quality of ontologies: Endurantism vs. perdurantism," *International Journal of Database Management Systems*, vol. 4, no. 3, June 2012.

[4] M. O'Connor and A. Das, "A method for representation and querying temporal information in OWL," in *Proceedings of Biomedical Engineering Systems and Technologies BIOSTEC 2010*, 2010, pp. 97–110.

[5] J. Hobbs and F. Pan, "Time ontology in OWL," (Online) http://www.w3.org/TR/owl-time/, (Accessed on November 2012).

[6] C. Gutierrez, A. Hurtado, and A. Vaisman, "Introducing time into RDF," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, pp. 207–218, February 2007.

[7] P. Hayes, "RDF semantics, W3C Recomendation, 10 february 2004," (Online) http://www.w3.org/TR/rdf-mt/, 2004, (Accessed on November 2012).

[8] M. Klein and D. Fensel, "Ontology versioning on the Semantic Web," in *Proceedings of the First International Semantic Web Working Symposium SWWS'01*, July 2001, pp. 75–91.

[9] S. Batsakis and E. Petrakis, "SOWL: Spatio-temporal representation reasoning and querying over the semantic web," in *Proceedings of the 6th. International Conference on Semantic Systems I–SEMANTICS 2010*, September 2010, pp. 15:1–15:9.

[10] ——, "SOWL: a framework for handling spatio-temporal information in OWL2.0," *Rule Based Reasoning, Programming, and Applications Lecture Notes in Computer Science*, vol. 6826, pp. 242–249, 2011.

[11] M. Koubarakis and K. Kyzirakos, "Modeling and querying metadata in the semantic sensor web: The model stRDF and the query language stSPARQL," in *The Semantic Web: Research and Applications*, ser. Lecture Notes in Computer Science, L. Aroyo, G. Antoniou, E. Hyvnen, A. Teije, H. Stuckenschmidt, L. Cabral, and T. Tudorache, Eds. Springer Berlin Heidelberg, 2010, vol. 6088, pp. 425–439. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13486-9_29

[12] K. Ryu and Y. Ahn, "Application of moving objects and spatiotemporal reasoning," 2001, a TIMECENTER Technical Report.

[13] A. Karmacharya, C. Cruz, F. Boochs, and F. Marzani, "Integration of spatial processing and knowledge processing through the semantic web stack," in *GeoSpatial Semantics*, ser. Lecture Notes in Computer Science, C. Claramunt, S. Levashkin, and M. Bertolotto, Eds. Springer Berlin Heidelberg, 2011, vol. 6631, pp. 200–216. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-20630-6_13

[14] J. Allen, "Maintaining knowledge about temporal intervals," *Communications of the ACM*, vol. 26, no. 11, pp. 832–843, Nov. 1983. [Online]. Available: http://doi.acm.org/10.1145/182.358434

[15] C. Strobl, *Encyclopedia of GIS Springer*. Springer, 2008, ch. Dimensionality Extended Nine–Intersection Model (DE–9IM), pp. 240–245.

[16] N. Brisaboa, I. Mirbel, and B. Pernici, "Constraints in spatio-temporal databases: A proposal for classification," in *Proceedings of the 3th. International Workshop on Evaluation of Modeling Methods in System Analysis and Design*, 1998.

[17] A. Artale and E. Franconi, "A temporal description logic for reasoning about actions and plans," *Journal of Artificial Intelligence Research*, vol. 9, pp. 463–506, 1998.

[18] R. Lake, "Introduction to GML," 2004, http://www.w3.org/Mobile/posdep/GMLIntroduction.html.


 ISBN: 978-1-61208-251-6 

# Generation of a Weighted Network Graph based-on Hybrid Spatial Data

Markus Prossegger
Dept. of Network- and Communication Engineering
Carinthia University of Applied Sciences
Klagenfurt, Austria
m.prossegger@cuas.at

***Abstract*—State-of-the-art network simulation and optimization techniques rank among the most studied problems in the field of operations research. While the mathematical models are studied in detail and nearly each network optimization problem has its already known solution in form of an optimal or heuristic algorithm, the underlying spatial data are the one key factor with respect to the optimization results. This paper examines the generation of weighted network graphs based on heterogeneous spatial data. Based on a general format, the normalized geobasisdata, an initial graph, is constructed. This graph is then used as input into our rule-based system to select and weight the edges to be in the final graph. The successful reduction of the complexity of the generated graph is shown in the experiments.**

***Keywords*—geographic information; spatial data; mathematical optimization; simulation; network construction; graph theory.***

## I. INTRODUCTION

Mathematical models of state-of-the-art network optimization and simulation techniques are based on network graphs consisting of vertices (i.e., points-of-interest) and pair wise joining edges between them. This work focuses on the missing link between the real world and the mathematical modeling - the weighted network graph based on spatial data. The graph is generated using spatial polygon data describing the land use of an area on the one hand, and spatial line- and point-objects, describing existing infrastructure on the other hand. Based on this spatial data originating from a number of hybrid sources, a rule-based expert system is used to construct a network graph as vital input in subsequent mathematical models. We focus on optimization models within the scope of telecommunication network construction. The models are intended to minimize network construction costs (including underground work and cable laying costs) as well as to maximize the number of customers that can be connected to the communication network infrastructure. An instance of such a model using weighted graphs is the simulation and optimization engine of fiber optic communication networks described in [1]. Here, the land use polygons are used to generate a graph originating from a cost raster, which describes the underground construction costs. While using a cost raster is a feasible way to generate a network graph representing network construction costs, a more sophisticated approach is needed to generate graphs by taking into consideration all kinds of real world information and being able to be computed in a reasonable time.

The present paper is divided into a preliminaries section, the section dedicated to the definition of normalized geobasisdata, details of our approach and experimental results followed by the conclusion.

## II. PRELIMINARIES

The subsequent simulation and optimization algorithms require undirected graphs as the fundamental data structure. The graph $G = (V, E, d)$ consists of $n = |V|$ vertices and $m = |E|$ edges. The distance of an edge $e_{ij} \in E$ connecting the two vertices $i \in V$ and $j \in V$ is given as a cost function $d_{ij} : E \to \mathbb{R}$. When calculating the shortest path or the minimum Steiner Tree [2], [3] as typical routing problems, the distance $d$ can be the Euclidean distance. The more sophisticated algorithms use travel time as the distance between two vertices. In case of scenarios considering the build-up of wired networks, the distances have to be construction costs, such as underground work or the costs for building cable poles.

The two-dimensional geographic data originate from hybrid sources, thus these data need to be prepared to serve as the basis for the construction of network graphs. There are three main sources for the geographical data:

(a) The geographic information system (GIS).
(b) The network information system (NIS).
(c) The Austrian digital cadastral map (DKM).

Each of the above listed items is needed for the construction of a consistent data source.

### A. GIS

In our case, the geographic information system includes typical information used in marketing scenarios and strategic decisions. It is important to know where the potential customers are located. The data showing the population density or the number of households collected in a population census are incorporated in the GIS as well. The GIS contains statistical data aggregated from public sources together with information gathered by the prosecuting company itself. The most important information for a network construction company is the information about the location of potential customers, the expected benefit, and the classification as private or public.

 ISBN: 978-1-61208-251-6 

The network operator knows the exact location and the return on investment of the current customers, but not for potential customers. The marketing division uses market surveys and other statistical data to predict the location of potential customers and the likely yearly sales.

### B. NIS

The network information system contains the information regarding all hardware components of the communication network as well as all logical links between these components. Typically, the NIS contains the most important business secrets of a network operating company. The following gives a list of the typical content of a NIS:

- Current and former customers.
- Network components.
- Physical cabling plan.
- Logical cabling plan.

### C. DKM

The Austrian Digital Cadastral Map is part of the official boundaries cadastre, which is the binding evidence of all parcel's boundaries. The DKM contains all public and private property and is available nationwide. Furthermore, it documents the type of land use of each parcel as well as buildings. Similar information is held in layers and together they form the DKM (a comprehensive interface description can be found in [4]):

- Boundaries.
- Parcel numbers.
- Types of land use (building land, forest, running water, standing water, etc.).
- Buildings.
- Control and Boundary points.

Formerly available only as an analog hard copy, the DKM was not only digitized but also enhanced using other official sources like Orthophotos and partition plans. Due to this reason, the quality of the digital map exceeds the quality of the analog version but it may include historical failures as well. There is a list of papers describing the aspects of spatial data quality [5], [6], [7], [8] as well as an ISO Standard regarding the quality of spatial data [9].

While the spatial accuracy is acceptable in most of mathematical simulation and optimization scenarios, the topological quality of the input data has to be ensured. There is some work proposed to identify spatial inconsistencies and incorrect object classifications using either manually defined spatial integrity constraints [10], [11], [12] or an automatic and incremental approach using decision trees proposed in [13] and improved in [14].

The Austrian DKM is used as one of the basic input into our approach and has to be normalized together with the other spatial input data.

## III. NORMALIZED GEOBASISDATA

The subsequent graph generation is designed as a completely automated process without the need of any user interaction. Due to this fact, the input data are stored in a predefined digital map format and the spatial objects must meet a set of conditions. In our approach, we have decided to use the ESRI Shapefile [15] as the digital map format. This open format is widely used and supported and stores spatial geometry and attributes as elements representing points, lines, and polygons.

The Normalized Geobasisdata (NGB) format [16] is an add-on to the ESRI Shapefile specifying the minimum qualitative and logical requirements of the spatial objects. It was developed in order to allow the automated generation of weighted network graphs based on any two-dimensional spatial data that represent surface data (i.e., land uses) in form of polygons at least. If the hybrid spatial data fulfill the specified NGB format, they qualify as input into the graph generation process.

The NGB format was originally developed for a simulation model dealing with the layout planning of a fiber-optics communication network in the year 2009. Since then it has been continuously adapted to the specific requirements of individual projects.

The majority of mathematical simulation and optimization models dealing with cable infrastructure planning or routing in general rely on spatial data as the main input source to stay real world compatible. We have defined the following list of objects as the required spatial information to be covered in the NGB format:

TABLE I
NGB OBJECT TYPES AND THEIR SPATIAL REPRESENTATION.

| Id | Object class | Spatial representation |
|---|---|---|
| a | Project area | Polygon |
| b | Land use | Polygon |
| c | Usable (own or third-party) infrastructure | Polyline |
| d | Infrastructure points | Point |
| e | Access points | Point |

The polygon describing the project area (a) as well as each polygon describing the land use (b) must show the following characteristics (in addition to some mandatory attributes described below):

- Valid and closed polygon.
- No crossing edges.
- Degree-two vertices only.
- No overlapping or equality with other polygons (see Figure 1).

A polyline referred to as usable infrastructure (c) can be associated with own infrastructure (e.g., the copper cabling in the access domain) or third-party infrastructure (e.g., leased lines). The spatial object must be a valid polyline. In case of any attributive restriction in the accessibility, there must be at least two infrastructure points assigned; hence the infrastructure can only be accessed via one of the two points (an open line infrastructure can be accessed at the masts only).

The infrastructure points (d) are attributive assigned to exactly one polyline of the usable infrastructure and represent any type of infrastructure objects that can be localized on exactly one position (e.g., shafts to access pipes, masts, hardware like splitters or routers, etc.).

 ISBN: 978-1-61208-251-6 

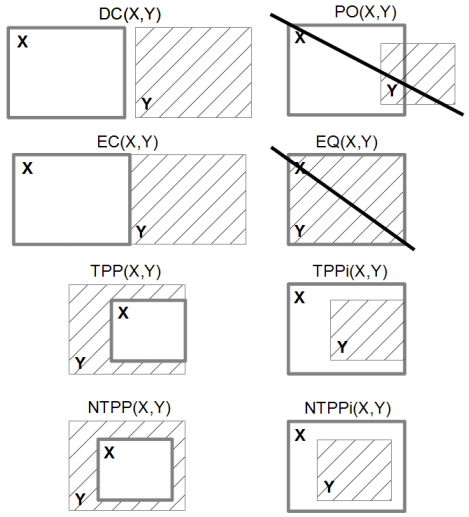
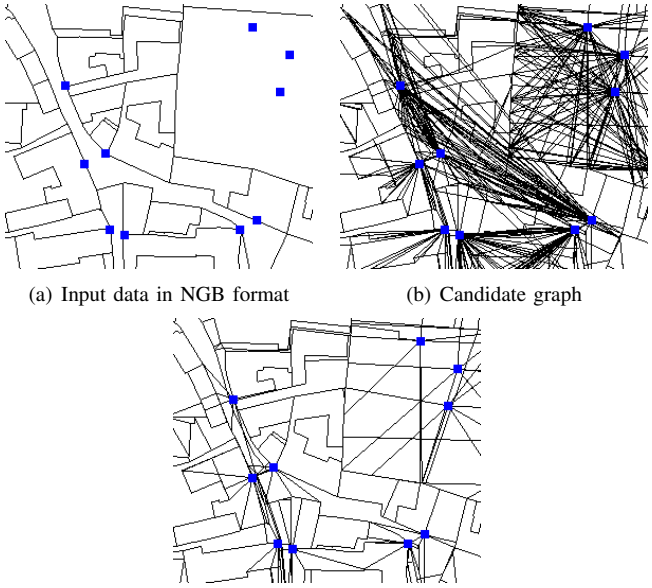



Fig. 1. All existing relations between pairwise polygons. EQ (equal) and PO (partly overlapping) relations are not allowed in NGB; Based on Region Connection Calculus (RCC-8) [17].



(a) Input data in NGB format (b) Candidate graph

(c) Generated graph

Fig. 2. (a) Preprocessed and enhanced input data, (b) the candidate graph according to Algorithm 1, and (c) the resulting generated graph . Access points are shown as squares.

The last object class is the class of access points (e). These are points representing the terminals in a following optimization. We distinguish between existing access points that are currently supplied by one or a group of connected usable infrastructure polylines, and potential access points that are not yet connected.

The presented approach of the graph generation makes use of geometric algorithms and algorithms from operations research. Thus, ambiguous relations or even holes between spatial objects cannot be allowed. This distinguishes between common geodatabase systems and spatial data in the NGB format, because there are no tolerances allowed.

Two points meant to be on the same position have to share the same coordinates. Furthermore, there are general specifications which cover the notation, the coordinate system, the locale, default attribute values, and a list of common abbreviations.

The next section describes our graph generation approach, that is based on spatial input data in NGB format.

## IV. Graph Generation

The process that we follow to generate a weighted network graph consists of three consecutive stages (see Figure 2 for the individual results):

- NGB preprocessing and enhancement.
- Generation of the candidate graph.
- Running the rule-based system.

In the following section, further details to the stages will be given.

### A. NGB preprocessing and enhancement

The preprocessing and enhancement of the input data are fully automated processes. As long as the input data fulfill the requirements in the NGB format, the generation of a weighted network graph will succeed. Moreover, the quality and usability of the generated graph are crucial in terms of topological errors within the spatial data. Furthermore, the succeeding process of assigning the correct weights to all edges in the graph is sensitive to the correct spatial classification.

To ensure high quality in the resulting graph, the input data are validated running the decision tree approaches described in [13] and [14]. Based on error free spatial data covering provincial, rural, suburban, and urban areas, a representative decision tree was constructed. Both approaches use this decision tree to validate the input data. The process will output warnings in case of topological errors and reclassify spatial objects according to the decision tree.

The validation is followed by the enhancement of the input data. Since underground work in crossroads areas and the subsequent obstruction in traffic should be avoided, the roadways are supplemented by polygons representing crossroads areas. Each crossing of at least two center lines of street polygons is identified and replaced by a polygon classified as crossroad area. The process runs automatically and produces suitable results (see Figure 3).

### B. Generation of the candidate graph

The goal of any graph-based mathematical optimization is to connect the access points to a given or new access network. As a result of the graph generation, each spatial object will be connected with other objects in the candidate graph. Algorithm 1 describes the construction of the initial candidate graph that is used as an input into the rule-based system.

 ISBN: 978-1-61208-251-6 

(a) (b)

Fig. 3. (a) Original NGB street polygons and (b) enhanced by crossroads area polygons.

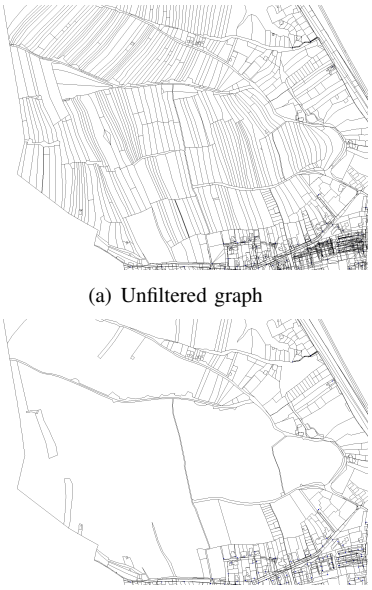
**Algorithm 1** : Generation of the candidate graph

```
Import all polygons P.
Import all infrastructure polylines L.
Import all infrastructure points I.
Import all access points A.
Import specifications regarding additional crossings.
for all polygons p ∈ P do
    Create edges representing the border of p.
    if p encloses an access point a ∈ A then
        Create (orthogonal) projections from a to p.
    end if
    if p should be enhanced with crossings then
        Create a crossing all x meter.
    end if
end for
for all polylines l ∈ L do
    if l has restricted access at two or more points i_{1..n}
    then
        Create edges between the points i_{1..n} representing l.
    else
        Create edges from the polyline l.
    end if
end for
for all infrastructure points i ∈ I do
    if i hits any created edge e then
        Split e into e_1 and e_2 at location i.
    else
        Create (orthogonal) projections to connect i.
    end if
end for
```

### C. Running the rule-based system

The rule-based system is based on the expert knowledge of network constructors. It is applied to the generated candidate graph to test for qualified edges. Each of the following questions will be answered with $yes$ or $no$ and determine the appearance of the edge in the final graph (edge is rejected, if all answered with $no$):

1) The edge is needed to ensure a connected graph.
2) The edge is part of the existing infrastructure.
3) The edge is not part of any land use to be filtered (compare Figure 4).
4) The edge is part of the best (least costly) possibility to connect spatial objects.



(a) Unfiltered graph

(b) Filtered graph

Fig. 4. (a) Unfiltered graph and (b) filtered by agricultural land uses.

While for the first three questions the answer is quite easy to find, the last question requires a combination of a list of rules with an algorithm from the area of combinatorial optimization to generate a valid answer.

After identifying all edges that are members in the final graph, the rule-based system is used to weight the edges. Depending on the originating land use or infrastructure, weights representing the network construction costs (or network usage costs in case of infrastructure) are assigned.

In the next section, we will describe the outcome of experiments we ran using input data from four different classification areas.

## V. Experimental Results

After running a series of experiments, the results are not only promising, but are proven to be valid in case of simulation and optimization scenarios in the field of the construction of hybrid communication networks.

The three main factors determining the quality of the generated graphs are:

(a) The real-world correlation.
(b) The correct cost assignment.
(c) The usability.

The (a) correlation and (b) weighting can only be tested by a comparison of a manually planned scenario with the result of an optimization using the graph.

The (c) usability is primarily defined by the focused optimization algorithm. The number of vertices and edges in the weighted graph must be small enough to allow the application of heuristics or optimal algorithms but at the same time it also must be large enough to be non-restrictive.

 ISBN: 978-1-61208-251-6 

In our experiments, we selected input data from four different classification areas:

- Urban;
- Suburban;
- Rural, and
- Provincial.

Table II provides the classification of the exemplary selected input data together with the dimensions.

TABLE II
SELECTED CLASSIFICATIONS AND DIMENSION OF ENCLOSED OBJECTS.

| Area | # Polygon Objects | # Line Objects | # Point Objects |
|---|---|---|---|
| urban | 20,569 | 59,408 | 33,090 |
| suburban | 18,497 | 30,997 | 16,361 |
| rural | 5,255 | 21,850 | 11,800 |
| provincial | 792 | 1,076 | 508 |

The results of applying our approach to the experimental data can be seen in Table III. The number of edges and vertices of the resulting graph is compared with a graph assumed to be fully connected (using all vertices of the spatial data objects). A fully connected graph represents the best possible real-world correlation (because nearly every path is present) but the least quality with respect to the applicability. Tests on small cut-outs of the experimental areas have shown that the optimization result using the generated graph equals the result using the fully connected graph.

TABLE III
DIMENSIONS OF THE GENERATED GRAPHS.

| Area | Fully connected graph | | Generated graph | |
|---|---|---|---|---|
| | # Edges | # Vertices | # Edges | # Vertices |
| urban | $9{,}384^6$ | $137^3$ | $346^3$ | $228^3$ |
| suburban | $6{,}850^6$ | $117^3$ | $263^3$ | $176^3$ |
| rural | $870^6$ | $41.7^3$ | $139^3$ | $88^3$ |
| provincial | $88.4^6$ | $13.3^3$ | $17^3$ | $14^3$ |

To evaluate the performance of the graph generation process, we measured the duration of the described approach. The application was realized using programming languages from the .NET Framework. Each algorithm, the user interface, as well as the control logic was implemented in the object-oriented programming language C#. The described rule-based system was realized using the functional programming language F#. The experiments were carried out on a standard personal computer with the following setup:

- OS: Microsoft Windows Server 64bit.
- CPU: 3.5GHz Dual Core.
- RAM: 8 GByte.

TABLE IV
PERFORMANCE OF THE APPROACH.

| Area | Duration in seconds |
|---|---|
| urban | 3,600 |
| suburban | 2,043 |
| rural | 882 |
| provincial | 519 |

The performance of the graph generation process can be seen in Table IV. As expected, the duration corresponds to the number of spatial objects enclosed in the area. For the simple reason that the graph needs to be generated only once, the duration is reasonable and acceptable.

## VI. CONCLUSION

In this paper, an approach for the generation and weighting of a network graph has been proposed. Introducing a normalization format called NGB to support a fully automated process of generating graphs using a wide range of hybrid spatial data as the input.

The experiments show that the approach is effective and efficient and that the weighted graph can be used as basic input into mathematical optimization algorithms. As a future investigation, we intend to explore ways to reduce the time needed for the graph generation process.

## REFERENCES

[1] P. Bachhiesl, M. Prossegger, H. Stoegner, J. Werner, and G. Paulus, "Cost optimal implementation of fiber optic networks in the access net domain," in *International Conference on Computing, Communications and Control Technologies*, 2004, pp. 334–349.

[2] A. Ivanov and A. Tuzhelin, *Minimal Networks: The Steiner Problem and its Generalizations*. CRC Press, 1994.

[3] R. D. Hwang, F. and P. Winter, *The Steiner Tree Problem*. North-Holland, 1992.

[4] "Katastralmappe SHP Schnittstellenbeschreibung, Version 2.0.1." BEV - Bundesamt fuer Eich- und Vermessungswesen, 2012.

[5] R. Wang and S. D.M., "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, pp. 5–34, 1996.

[6] L. Leo Pipino, Y. W. Lee, and R. Y. Wang, "Data quality assessment," *Commun. ACM*, vol. 45, no. 4, pp. 211–218, 2002.

[7] B. K. Kahn, D. M. Strong, and R. Y. Wang, "Information quality benchmarks: product and service performance," *Commun. ACM*, vol. 45, no. 4, pp. 184–192, Apr. 2002.

[8] A. Jakobsson and F. Vauglin, "Status of data quality in european national mapping agencies," in *Proceedings of the 20th International Cartographic Conference*, vol. 4, 2001, pp. 2875–2883.

[9] "19113 Geographic Information - Quality Principles," in *ISO/TC 211*. International Organization for Standardization (ISO), 2002.

[10] T. Ubeda and M. J. Egenhofer, "Topological error correcting in gis," in *Proceedings of the 5th International Symposium on Advances in Spatial Databases*, ser. SSD '97. London, UK, UK: Springer-Verlag, 1997, pp. 283–297.

[11] K. A. V. Borges, C. A. Davis, Jr., and A. H. F. Laender, "Database integrity," J. H. Doorn and L. C. Rivero, Eds. Hershey, PA, USA: IGI Publishing, 2002, ch. Integrity constraints in spatial databases, pp. 144–171.

[12] M. Mostafavi, G. Edwards, and R. Jeansoulin, "An ontology-based method for quality assessment of spatial data bases," in *Proceedings fo the Third International Symposium on Spatial Data Quality*, vol. 28, 2004, pp. 49–66.

[13] M. Prossegger and A. Bouchachia, "Incremental identification of topological errors in spatial data," in *The 17th International Conference on Geoinformatics*, Aug. 2009, pp. 1–6.

[14] M. Prossegger and A. Bouchachia, "Incremental semi-automatic correction of misclassified spatial objects," in *Proceedings of the Second international conference on Adaptive and intelligent systems*, ser. ICAIS'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 16–25.

[15] "ESRI Shapefile Technical Description," in *An ESRI White Paper*, July 1998.

[16] M. Prossegger, "Normalized Geobasisdata (NGB) - technical requirements v.3.0," FHplus Project Netquest, Carinthia University of Applied Sciences, Tech. Rep., October 2012.

[17] A. David, Z. Cui, and A. Cohn, "A spatial logic based on regions and connection," in *Proc. KR-92*, 1992, pp. 165–176.

 ISBN: 978-1-61208-251-6 

# Trajectory Data Mining: a Novel Distance Measure

Ayman Al-Serafi
Business Intelligence & Enterprise Data Warehousing
Teradata Corporation
E-mail: ayman.al-serafi@teradata.com

Ahmed Elragal
Department of Information Systems
German University in Cairo, GUC
E-mail: ahmed.elragal@guc.edu.eg

***Abstract*— There is currently an increasing availability of large spatiotemporal datasets. Sequences of spatiotemporal data or paths, also known as trajectories, can be captured by modern technology and stored in moving-object databases (MOD) or a trajectory data warehouse. It is a common challenge for knowledge discovery within MODs to query proximities and distances, e.g. in clustering trajectories. Previously adopted distance measures focus on the complexity of geometric and/or mathematical models of trajectories, while ignoring several aspects common to all spatiotemporal trajectories, e.g. direction, distance covered, and duration. This research introduces a more comprehensive approach for trajectory distance measurement in spatiotemporal applications. The approach is simplified, yet novel, introducing a new set of dimensional variables, therefore called the Multi-Dimensional Trajectory Distance Measure (MTDM). The accuracy and relevance of MTDM is evaluated in experiments using multiple proximity metrics, for example MTDM based on Euclidean proximity calculation. A geospatial data analysis framework is utilized in the experiments. Efficiency evaluation of MTDM showed the feasibility of applying the measure to various trajectory datasets.**

***Keywords-distance measurement; geospatial data analysis; moving-object databases; spatiotemporal data; trajectory data mining.***

## I. INTRODUCTION

There is currently an increasing availability of large spatiotemporal datasets. Data about movements and trajectories of objects are commonly captured using technology like Global Positioning System (GPS), Radio Frequency Identification (RFID) and Global System for Mobile communications (GSM) [8, 17]. The trend of increasing spatiotemporal data has been also supported by the advances in database management systems (DBMS) which support such kind of data. This can be seen with the increasing adoption of geospatial and temporal capabilities in DBMS which can support Mobile Objects Databases (MOD) [10, 14, 15].

Spatiotemporal data consists of a collection of points which have location and temporal references. Temporal references of spatiotemporal data are commonly denoted by a date and an instance on the world time system (like GMT). Meanwhile, location references from spatiotemporal points are commonly stored as spatial references (X and Y references on the Cartesian coordinate system), or as geospatial references consisting of a location on Earth referenced by a geodetic system like the World Geodetic System (standard format for GPS location references).

A trajectory is a path consisting of an ordered set of spatiotemporal points [8, 16]. This can be defined for any trajectory 'T' which can be seen as an ordered set of spatiotemporal points consisting of 3 dimensions: location in terms of x-coordinate 'x', y-coordinate 'y' and temporal dimension in terms of time 't'. This is formally defined as $T = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_n, y_n, t_n)\}$ and can be seen in Figure 1.

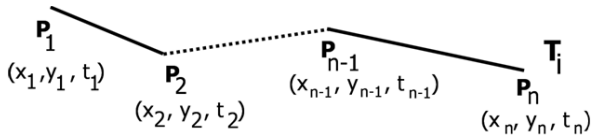



Figure 1. Formal representation of a trajectory consisting of 'n' points.

After storing trajectories in MODs, or a trajectory data warehouse (TDW), data mining is frequently used to analyse and to find knowledge within this data. Data mining involves multiple techniques like clustering, association rules and outlier detection. For techniques like clustering, proximity measurements are frequently used [8, 13, 16]. Proximity measurements can be also used to find outliers in datasets [4]. Examples of outliers include fraudulent transactions in e-commerce. Proximity measurements can also be used for classification using the nearest neighbour techniques [18] and trajectory similarity [14].

The aim of this paper is to develop a novel measurement of distance (proximity) between trajectories for spatiotemporal applications, and to test this distance measurement within a geospatial data analysis framework. The paper is organised as follows: in Section II a background about the research problem and proximity measures is provided, in Section III the Multi-Dimensional Trajectory Distance Measure (MTDM) is introduced, Section IV analyses the efficiency of our algorithms, Section V tests our approach in a series of experiments utilising the geospatial framework followed by a discussion in Section VI, and finally Section VII concludes the research study and advises on future research.

## II. PROBLEM STATEMENT

Proximity (or similarity) can be defined as the degree of how close, or alike, are two instances [4]. Similarity coefficients measure the relationship between two individual items based on a number of variables. Distance, on the other hand, is seen as the degree of differences between instances based on specific dimensions or variables [16]. Different proximity metrics exist to calculate the distances between

 ISBN: 978-1-61208-251-6 

different instances under study. This includes Euclidean [12, p. 21], City Block [12, p. 21], Canberra [12, p. 21], and Chebyshev [11].

When calculating proximity between all instances the result is summarised in a proximity matrix [4]. The proximity matrix is a symmetric matrix with the diagonal equal to 0 when it is a distance measure. The number of unique elements for the matrix 'N' (which need to be calculated on one side from the diagonal; the other side is neglected as it is a mirror of the same values -similar to any symmetric matrix) is given by equation 1, where 'n' is the number of instances [4]. The proximity matrix compares all instances with each other. For distances the proximity matrix of trajectories is defined by the matrix in Figure 2.

$$N = \frac{n(n-1)}{2} \quad (1)$$

Previously researched measurements of trajectory distances include: the Euclidean distance, the Longest Common Sub-Sequences (LCSS) distance, Dynamic Time Warping (DTW) and edit distances like real sequences (EDR) and real penalty (EDP) [6, 8, 9, 13, 15].

$$\begin{bmatrix} & T_1 & T_2 & T_3 & \cdots & T_n \\ T_1 & 0 & & & & \\ T_2 & d_{(2,1)} & 0 & & & \\ T_3 & d_{(3,1)} & d_{(3,2)} & 0 & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \\ T_n & d_{(n,1)} & d_{(n,2)} & d_{(n,3)} & \dots & 0 \end{bmatrix}$$

Figure 2. Trajectory distance proximity matrix.

Previously researched measurements focus on the time-series approaches and geometric complexities similar to [5]. These measurements consider the trajectory as a time-series of spatiotemporal points which puts the burden of some limitations like time-shift complexities (related to trajectory distance measurement at different segments, or sub-trajectories, and for different temporal windows between the trajectories segments compared). Another limitation can be the duration of the trajectories to be compared. If they are not equal it also puts some complexities on the algorithms. This can be seen in [5] where the efficiency of such algorithms for non-equal time-windows can have high complexities.

Simple Euclidean distance is also a frequently utilised distance function. An example can be seen in the MINDIST Euclidean function used for trajectory queries in [10]. Distance measures like DTW and simple Euclidean distance do not allow for some points in a trajectory to be unmatched, however other approaches like LCSS allow for this [9]. On the other hand, algorithms based on Euclidean approaches similar to that of [19] require trajectories to have equal length.

Edit distance algorithms like EDP measure the cost of transforming one of the compared trajectories to the other. EDP utilises real spatial distances [13]. Both of EDR and LCSS focus on the matching or non-matching characteristic of comparison points in two trajectories. A match is considered so compared to a user-defined threshold for the distance measurement [13]. EDR returns the number of transformations required to match the trajectories being compared [13]. The LCSS algorithm also returns a non-metric distance measurement [17].

Other unique distance measures were also proposed. For example, spatial distance between trajectories can be a function of the area of the two-dimensional regions between trajectory intersections. This approach is called Locality In-between Polylines (LIP) [15]. This has an advantage that it calculates sub-sequence similarity between trajectories. However, the approach is limited by assumptions of directional similarities and intersections between trajectories requiring additional workarounds [15].

Other trajectory distance measures also include [8] which allow calculating distances for non-overlapping trajectories in terms of temporal aspects. Their approach is a modified Euclidean distance of the spatial dimensions, where the modifications consist of temporal aspects of the trajectories. However, their approach only considers spatiotemporal aspects related to privacy preservation. The work of [9] integrates temporal aspects of the trajectories compared into the distance measurement too. The shape of trajectories was also the focal interest of approaches like [19].

The majority of the state-of-the-art approaches consider (dis)similarities of geometric movements and proximity. The focus of those approaches is on the complexity of geometric and/or mathematical models of trajectories, which ignores several aspects common to all spatiotemporal applications like total distance covered by trajectories, duration of trajectories, minimum/maximum distances between trajectories, etc. In our research, we tackle the problem of finding trajectory proximities in geospatial space and taking the overall characteristics of trajectories into account. This approach was recommended for future research by multiple research studies like [5] which recommend using non-geometric and non-time-series approaches to find proximity between trajectories.

## III. THE MULTI-DIMENSIONAL TRAJECTORY DISTANCE MEASURE (MTDM)

Most of the trajectory distance measurements analysed trajectory similarities based on matching sub-sequences of the trajectories in a manner similar to time-series mining approaches. Such approaches assume that trajectories are a series of fluctuations in space and over time. For example, [5] uses average Euclidean distance at specific time-windows to find proximity between trajectories. [7] can be referred to for a more detailed discussion about the time-series approaches. In addition, many of those approaches only consider sub-trajectories (segments of trajectories) instead of complete trajectories [5].

We argue that attention for the sub-sequence matching approaches from the time-series literature and which focus on the geometric (shapes) and mathematical (time-series theories) aspects neglect the specific properties of moving-object trajectories which are relevant to spatiotemporal applications. This includes properties like the duration of movement, the direction of movement, the locations of

 ISBN: 978-1-61208-251-6 

movements, etc. Therefore our approach will adopt the latter multidimensional approach examining moving-object trajectories. This was previously introduced by similar research studies which adopt the concept of applying trajectory properties to distance measures, for example [15] which introduced the speed and direction dimensions.

In our distance measurement we will apply this approach which uses distance metrics, like Euclidean and City Block, and which can handle the different data types of trajectory data. Multiple dimensions and factors of trajectories were recommended for computing proximities, such as spatial distance, trajectories' start and end points proximities, geometric movement patterns, and movement dynamics (like speed) [15, 16]. Such combination and special considerations was recommended by [16] and will therefore be considered in our approach.

The aim of our approach is to measure the distance between all trajectories in an MOD via an approximate, yet accurate, approach which focuses on the characteristics of trajectories. Approximation can be seen, for example, in that our approach does not consider the exact shape of the trajectories (only considers the overall occurrences of directionality, like northbound movement occurrences, relaxing the geometric assumptions) and the sequence of points does not matter too (relaxing the assumptions of time-series theories).

Approximation of trajectory distances is not a new concept and has been studied in previous research. For example, similar directional approximation based on the geographic orientations was also utilised in the study of [15]. Such approximation can improve the efficiency of the distance operators utilised in the algorithms without neglecting parts of the whole trajectory (similar to methods related to sub-trajectory distance measurements, like [5], which only seeks to find distance between trajectory segments and not the whole trajectory and which have some limitations concerning efficiency of the algorithms.) The sequence of points can be relaxed as the interest is not in the sequence patterns but rather on the complete trajectory proximity (with the shape dimension being captured by total directionality of the trajectory rather than the local directionality within trajectory segments or sequences).

Our approach will be called Multi-Dimensional Trajectory Distance Measure (MTDM) as it will involve a more comprehensive approach for trajectory distance measurement and which takes into account multiple characteristics of trajectories. MTDM can be defined for two trajectories $T_i$ and $T_j$ as a simplified function of the multiple dimensions as in equation 2. The MTDM distance function consists of multiple variables as described in Table 1.

$$MTDM_{T_i,T_j} = f(GeoDist, Distance_{\Delta}, Direction_{\Delta}, Duration_{\Delta}, Min\_GeoDist, Max\_GeoDist, Min\_ExtDist, Max\_ExtDist) \quad (2)$$

MTDM can be described as follows:-

1. Within each trajectory T, compare each point of the trajectory with the point succeeding it in order to calculate the direction of movement (see Figure 3), the duration of movement and the distance covered between both points. The output of this step is stored in a trajectory spatiotemporal table (see Trajectory_ST in the data model depicted in Figure 4.). In this table, each point in a trajectory is assigned an ID (ID_COL), an identifier of the object being traced (Object_id), identifier for the trajectory (trajectory_id), and the date, time and geospatial location (mov_Date, mov_time, geo_location respectively). The calculated distance of movement, duration of movement, and direction of movement are stored as geo_distance, duration and geo_direction respectively.

TABLE I. MTDM DIMENSIONS DESCRIPTION.

| MTDM Dimension | Description | Detailed Step(s) |
|---|---|---|
| **GeoDist** | Average geographic distance measurement between two trajectories. | 1. Find the smallest distance between each point of trajectory $T_i$ with all points in $T_j$ (and vice versa).<br>2. Average all the geographic distances calculated from step 1 and for each of the two trajectories.<br>3. *GeoDist* is finally assigned as the smallest average from the two calculated in step 2 |
| **Distance$_{\Delta}$** | The difference between the total distances covered between the two trajectories. | • Find the absolute difference between total distances covered by the two trajectories |
| **Direction$_{\Delta}$** | The difference in directional movements between the two trajectories | • Find the absolute difference in directional counts for each of the 8 directions: N/ NE /E / SE /S/ SW/ W/ NW |
| **Duration$_{\Delta}$** | The difference between the total durations between the two trajectories. | • Find the absolute difference between durations of the two trajectories |
| **Min_GeoDist** | The minimum geographic distances between points in $T_i$ and $T_j$ | • Find the smallest distance between any point of trajectory $T_i$ with any point in $T_j$ (and vice versa). |
| **Max_GeoDist** | The maximum geographic distances between points in $T_i$ and $T_j$ | • Find the largest distance between any point of trajectory $T_i$ with any point in $T_j$ (and vice versa). |
| **Min_ExtDist** | The minimum geographic distances between the extreme (first and last) points of the trajectories. | 1. Find the distance between the first and last point of trajectory $T_i$ with the first and last point in $T_j$ (and vice versa).<br>2. Assign Min_ExtDist as the smallest distance from step 1. |
| **Max_ExtDist** | The maximum geographic distances between the extreme (first and last) points of the trajectories | 1. Find the distance between the first and last point of trajectory $T_i$ with the first and last point in $T_j$ (and vice versa).<br>2. Assign Max_ExtDist as the largest distance from step 1. |

2. Calculate the sum of all distances and durations for each trajectory and store it into the Trajectory_Info table (see the data model in Figure 4) as distance and duration respectively. Count the number of occurrences for each direction and store it into the Trajectory_Info table (count the

directional orientation N/NE/SE/S/SW/W/NW for each trajectory.). The total number of points in the trajectory is also stored as "trajectory points number".

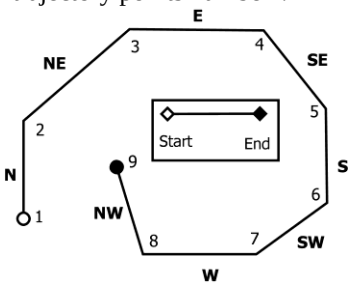



Figure 3. Calculating the direction of travel between each point and the point after it in a trajectory. For example, there is a positive change in both the x-axis and the y-axis between points 2 and 3 of the trajectory resulting in a direction of NE.

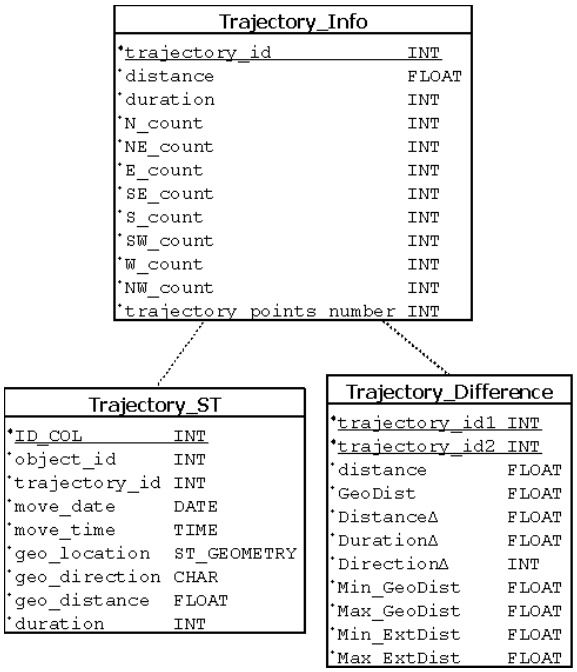



Figure 4. Trajectory MTDM data model.

3. Convert the direction counts of the trajectories (in table Trajectory_Info) into percentages by dividing the counts by the total number of points and multiplying by 100%. This normalises the directional counts relevant to the number of points in each trajectory so that trajectories with different number of points can be compared.

4. The differences between the trajectories' characteristics (described in Table 1) are then calculated and stored in the Trajectory_Difference table (Figure 4). Figure 5 depicts the Min_ExtDist and Max_ExtDist dimensions. The geographical distances between points of trajectory Ti and trajectory Tj is calculated using the geospatial function ST_SphericalDistance (which calculates spherical geospatial distance, see [2] for a description of the function).

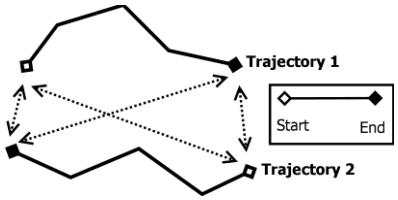



Figure 5. The first (trajectory start position) and last point (trajectory end position) of each trajectory is compared with the first and last point of the second trajectory. Four total comparisons (dotted-lines) are calculated.

5. Finally, standardise all the distance measurements '$X_i$' in the Trajectory_Difference table by rescaling them between 1 and 100 using equation 3. Standardization to the distance measures is done by rescaling in order to make sure that there is no single (large) dimension dominating the distance measurement. This procedure is recommended so that the distance between two trajectories is invariant to spatial scaling and shifting.

$$Standard(X_i) = 1 + \frac{\big((X_i - \min(X)) * (100 - 1)\big)}{(\max(X) - \min(X))} \quad (3)$$

After completing the MTDM procedures and calculating specific characteristics and proximities between trajectories, the overall proximity can be calculated using multiple proximity metrics. The five proximity metrics we are going to use are modified Euclidean, City Block, Canberra, Chebychev, and an average proximity metric. The average proximity metric will also consider all of the dimensions as shown in equation 4. Average proximity metrics, like average Euclidean distance, was used in other trajectory similarity measures too [5].

$$MTDM_{average} = \frac{\begin{array}{c} GeoDist + Distance_{\Delta} \\ + Direction_{\Delta} + Duration_{\Delta} + Min\_GeoDist \\ + Max\_GeoDist + Min\_ExtDist + Max\_ExtDist \end{array}}{8} \quad (4)$$

## IV. EFFICIENCY EVALUATION

MTDM can be decomposed into two main algorithms. *Algorithm 1* is used to calculate the distance, duration and direction between each point and the point after it for all points in each trajectory (step 1 of MTDM in Section III); and *Algorithm 2* is used to calculate the differences between all trajectories (step 4 of MTDM in Section III). Algorithm 1 runs in linear time $O(t \cdot n)$ where n is the number of points in each trajectory and t is the number of trajectories. Algorithm 2 runs in the polynomial time ( $\frac{t(t-1)}{2} \cdot n^2$ ). The efficiency of algorithm 2 is based on comparing all trajectories together as expressed by equation 1, and within the comparison of every pair of trajectories there are quadratic steps executed on the points from both trajectories as described in step 4 of the MTDM approach (including finding the minimum, maximum and average distances between the trajectories- in Section III).

The overall big-O-notation is therefore $O(n^2)$. The efficiency of MTDM is therefore comparable with other trajectory distance measures which run in $O(n^2)$ polynomial time (with the exception of simple Euclidean distance which runs in O(n) linear time under the assumption that

 ISBN: 978-1-61208-251-6 

trajectories compared have equal number of points). The efficiencies of other algorithms from the time-series literature are: $O(n^2)$ for both EDP and DTW [13]. LCSS and EDR are non-metric and compare all points in the trajectories resulting into an efficiency of $O(n^2)$ too [13].

## V. EXPERIMENTS

In order to evaluate MTDM and its effectiveness with spatiotemporal applications a number of experiments are conducted. The first experiment tests the effectiveness of MTDM on a real dataset of spatiotemporal trajectories. In addition, experiment 1 compares MTDM with the simple Euclidean distance measurement as a benchmark. The simple Euclidean distance was calculated for two trajectories $T_i$ and $T_j$ as described in equation 5. The 'x' and 'y' are the coordinates of the i'th point of the trajectory (or j'th point of trajectory 2) retrieved using the standard ST_X and ST_Y methods in spatial databases and applied on the geo_location of each point from the Trajectory_ST table (Figure 4) –see [2]. The distance and duration are retrieved from the Trajectory_Info table and are used with the simple Euclidean distance in order to make it comparable with the MTDM approach (which also includes the duration and distance covered for trajectories. The simple Euclidean approach must also include those factors in-order to keep the evaluation non-biased towards the approach with the more representable variables).

$$\sqrt{\sum_{i,j=1}^{n} \begin{matrix} (x_i - x_j)^2 + (y_i - y_j)^2 + (distance_i - distance_j)^2 \\ + (duration_i - duration_j)^2 \end{matrix}} \quad (5)$$

We will not compare our MTDM approach with other time-series based approaches (like DTW or LCSS) because the shapes and geometric matching of trajectories is relaxed as described earlier in this paper. The main focus therefore is to assess the degree of relevance of the MTDM dimensions for assessing dissimilarity of trajectories in spatiotemporal applications.

Another experiment was conducted to test the accuracy of the MTDM distance measure as compared to simple Euclidean distance in experiment 2. Experiment 2 introduces a new technique for evaluating proximity measures which better differentiates the different measures and metrics evaluated as compared to experiment 1 (which uses classical evaluation techniques from previous literature).

For both experiments, we have used a subset of the trajectories dataset which consisted of 145 spatiotemporal trajectory traces from 2 school buses (moving-objects) transporting students within in the vicinity of the Athens metropolitan area, Greece (See [10] for more details about the dataset). The dataset was retrieved from [3].

The approach used in the experiments to execute the distance measurement on the sample datasets consisted of multiple steps and tools as summarized in the framework diagram (see Figure 6). First, the data is loaded into the Teradata TDW using Extract, Load and Transform (ELT) techniques. The original data was in the WGS 84 projection utilised in GPS. This format and multiple other projections are supported with Teradata Geospatial. To load spatiotemporal maps/GIS data from GIS/Map formats into the TDW we have used the TDGeoImport tool. The tool automatically converts the different spatiotemporal formats into the standard Teradata ST_Geometry format stored in the Teradata MOD / TDW. This allows for Teradata Geospatial functions to be used on the spatiotemporal data, which can also be combined with non-spatial data allowing for deeper and richer analytics. This data can also be visualized using data visualization tools by using the GeoServer geospatial data integration tool [1] or by direct SQL-querying over the ST_Geometry fields.

The distance measurement calculation then takes place using the Teradata Stored Procedures. The programming language used in implementing the algorithms and distance functions was in standard Teradata Database Stored Procedures (V13.0).

Following this, visualization takes place on Google Earth. There are two alternatives to achieve this goal (and both techniques were used). The first alternative uses GeoServer [1] middleware between the Teradata TDW and the Google Earth interface. GeoServer uses the Web Map Service (WMS) protocol. WMS is a standard protocol for serving geo-referenced map images over the Internet that are generated by a map server using data from a GIS database. GeoServer supports the Teradata Database (MOD). In this situation, GeoServer acts as the geospatial middleware connecting the client (Google Earth application) to the maps or geospatial data stored in the TDW. Alternatively, the data to be visualized can be exported by the TDGeoExport tool. This allows exporting the data to the Google Earth KML format, which can then be visualized.

### A. Experiment 1: Classification accuracy

In the first experiment we test the alternative distance metrics applied within MTDM and compare it with the simple Euclidean distance. We use the "leave-one-out" approach which was used before in previous literature like [7] and [6]. In this approach the trajectories are already assigned to clusters and labelled. Using the distance measure to be evaluated, the nearest trajectory to each of the trajectories is found and the class label of the former trajectory is assigned to the latter trajectory. This is similar to the 1-nearest-neighbour classification approach which classifies an object to the same class-label as its nearest neighbour. If the assigned class matches the real (actual) class then it is a hit (accurate), otherwise it is a miss (inaccurate). The accuracy rate is then calculated as in equation 6.

For this experiment 78 trajectories were sampled from the original dataset and which consisted of 5850 points (75 points in each trajectory). The trajectories retrieved were selected so that they belong to one of 4 clusters (visualized in Figure 7): cluster 1 consisting of medium to long trajectories at areas remotely located from Athens city centre and shorter trajectories closer to the centre or NE area but with more jagged shapes (blue), cluster 2 consisting of trajectories of small to medium length trajectories with close proximity to

 ISBN: 978-1-61208-251-6 

Athens city centre but with less jagged shapes (red), cluster 3 consisting of trajectories lying NE of Athens city centre and of small to medium length trajectories (green), and finally cluster 4 consisting of all longer trajectories located in remote regions around Athens (purple).

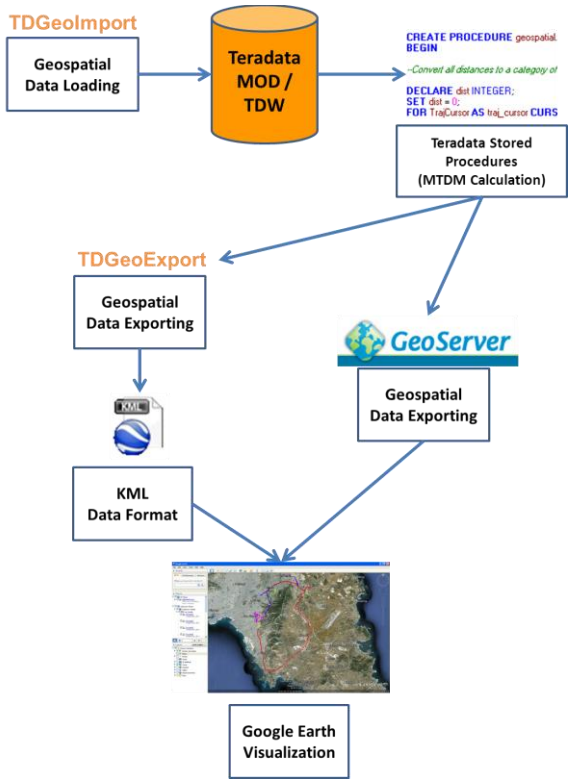



Figure 6. The Geospatial Data Analysis Architecture.

This experiment resulted into absolute accuracy of 100% for all measures including the MTDM implementations (Euclidean, City Block, Canberra, Chebyshev, and average proximity metrics) and the simple Euclidean distance. In order to distinguish between the different measures another experiment was conducted which evaluates the accuracy of the distance measures on domain-expert labelled data (See experiment 2).

$$Accuracy\ Rate = \frac{No.of\ hits}{No.of\ hits + No.of\ misses} \quad (6)$$

*B. Experiment 2: Apriori-knowledge validation*

In the second experiment we compare the accuracy of the different proximity metrics utilised within MTDM and compare it to the simple Euclidean distance measurement. For this experiment the original dataset was sampled for 1386 points making up all 11 trajectories from two days (9th and 12th of February 2001) and resulting into 55 trajectory comparisons (as calculated for 'n' trajectories by equation 1). The duration of a trajectory was considered as a calendar day (trajectories were divided into sub-trajectories covering a single day). Those trajectories are depicted in Figure 8 (each individual trajectory visualized with a unique colour).


Figure 7. Visualization of the 4 trajectories' clusters used in experiment 1.

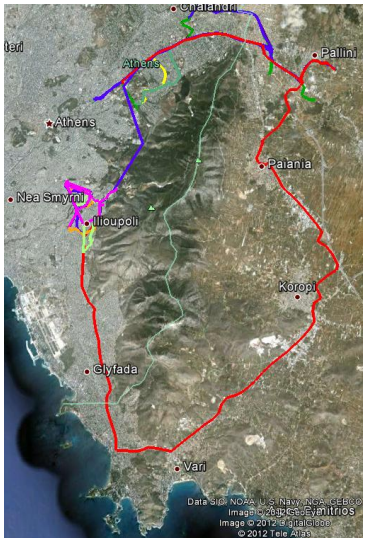



Figure 8. Visualization of the trajectories used in experiment 2.

We introduce a new approach for testing alternative distance measures based on expert apriori-knowledge evaluation. To test the accuracy of our approach, each of the distance measures was validated against a repository of pre-evaluated trajectory distance assigned by data mining experts. For our experiment we had 3 repositories of apriori-knowledge from 3 different data mining experts. The repositories identify the distance between two trajectories on a scale of four degrees of similarity {A, B, C, D}. Similarity of A would mean the two trajectories compared (from the expert's point-of-view) are most similar and similarity of D would mean least similarity. On the other hand, to convert the numerical MTDM / simple Euclidean distances 'd' to similar categories (as assigned in the knowledge repositories) we used the following conversions for the different ranges: ($1 \leq d \leq 25 \rightarrow$ A), ($26 \leq d \leq 50 \rightarrow$ B), ($51 \leq d \leq 75 \rightarrow$ C), ($76 \leq d < \infty \rightarrow$ D).

The analysis and interpretation of human analysts is recommended to use in order to capture the embedded meanings within the data [16]. Human cognition is suitable for such classification problems because human knowledge and common-sense are utilised to capture what seems to be

correct to human interpretation [16]. This apriori-knowledge testing-set will be used to evaluate the classification accuracy of MTDM using multiple alternative proximity metrics as compared to the simple Euclidean distance.

For each of the proximity measures we constructed a proximity matrix which compares all trajectories together similar to Figure 2. This totalled to 55 comparisons from the 11 trajectories as calculated by equation 1.The proximity measures were then converted to categories {A, B, C, D} similar to the approach used within the apriori-knowledge repositories. To validate the accuracy of each proximity measure with the benchmark of the human experts we constructed a confusion matrix which evaluates the degree of agreement in assigning (dis)similarity between two trajectories (see Figure 9). Full-agreement classifications are across the diagonal, namely AA, BB, CC, and DD. Those were assigned a coefficient of 100% accuracy. Partial agreement of a single degree difference (like classifying a distance between two trajectories as C instead of B) was assigned 66.67% classification accuracy. Two degree difference in classification agreement was assigned 33.33% accuracy while a three degree difference in classification agreement was assigned 0% accuracy.

MTDM utilising the alternative five proximity metrics was then evaluated along with the simple Euclidean distance. The resulting accuracy of the proximity measures (calculated for each proximity measure from its confusion matrix as in Figure 9) are summarised in Table 2 (in order of descending accuracy rates).

| | Computed Classification | | | |
|---|---|---|---|---|
| Actual Classification | $A$ | $B$ | $C$ | $D$ |
| $A$ | $100\% * AA$ | $66.67\% * AB$ | $33.33\% * AC$ | $0\% * AD$ |
| $B$ | $66.67\% * BA$ | $100\% * BB$ | $66.67\% * BC$ | $33.33\% * BD$ |
| $C$ | $33.33\% * CA$ | $66.67\% * CB$ | $100\% * CC$ | $66.67\% * CD$ |
| $D$ | $0\% * DA$ | $33.33\% * DB$ | $66.67\% * DC$ | $100\% * DD$ |

Figure 9. Confusion matrix for evaluating the trajectory distance measures in our experiments.

## VI. DISCUSSION

The In this paper, multiple proximity metrics and different combinations of dimensions from MTDM were tested in a spatiotemporal application. The results explain the capability of MTDM in accurately estimating the distance between trajectories. This was shown with its capability to correctly classify 100% of the instances in experiment 1 and by the overall accuracy rate ranging between 72.12% (with the MTDM Canberra metric) and 80.61% (with the MTDM Euclidean metric) in experiment 2. It must be noted that the top scoring proximity measures in experiment 2 were based on the MTDM approach which outperformed the simple Euclidean approach. The best MTDM distance metric was Euclidean as expected from the literature [7]. The average metric came second which can mean that all of the MTDM dimensions proposed had equal weights in the evaluation of trajectory distance and are all important as validated against the apriori-knowledge repositories. This was equally followed by the City Block MTDM implementation which had the same accuracy rate as the average MTDM implementation.

On the other hand, the Chebychev implementation of MTDM didn't perform as well as the Euclidean, City Block and average implementations because it takes the maximum distance making it highly sensitive to noise (or peaks in the variables) [7]. The Canberra implementation didn't perform as well as the other implementations of MTDM too. This is probably due to the high dimensionality of MTDM which can affect the Canberra distance. This is mainly because the Canberra distance assumes there is an origin (of expected values) and with values scattered around the origin. This was not the case for our trajectories experiment. In our experiments, the trajectories had random movements not scattered around a specific origin.

As for the simple Euclidean measurement, it came lagging behind all the MTDM implementations. This indicates the inadequacy of the simple Euclidean approach to estimate proximity of trajectories as human experts would expect. Alternatively, the MTDM implementations had considerable improvements in approximating the proximities of trajectories indicating the importance of the dimensions proposed in MTDM. Overall, the experiments indicate that the best proximity metric to use with MTDM is the Euclidean measure.

Concerning the correlation between the dimensions in MTDM, it was observed (using the dataset in experiment 1) that the absolute Pearson correlation coefficients between variables are negligible/small as they range between 0.0026 and 0.2192. The exception was the distance metrics relying on geospatial distance (GeoDist and ExtDist distances) which have high correlation ranging between 0.9536 and 0.9984 (as expected because they use the geospatial distance as a common base-metric). It was decided to keep all geospatial distance metrics for the experiments to maintain a comprehensive approach for MTDM, however, alternatively the variables relying on geospatial distance metrics can be used interchangeably according to application requirements.

There were some limitations with our approach and experiments. As for the approach, the approximation of the directional properties to the 8 compass directions could be considered as a limitation, however, we consider it an advantage which saves calculation costs of computing exact degree-based differences. Concerning the experiments, the accuracy rates were estimated but they could be largely affected by the reliability of the human approximations (for experiment 2).

## VII. CONCLUSION

This research paper introduced a more comprehensive approach for trajectory distance measurement in spatiotemporal applications called MTDM. The approach is simplified, yet novel, introducing a new set of dimensional variables. MTDM can be used in trajectory data mining techniques requiring trajectory proximity measurement like trajectory clustering or similar trajectory querying.

 ISBN: 978-1-61208-251-6 

TABLE II. THE ACCURACY RATES OF THE PROXIMITY MEASURES IN EXPERIMENT 2.

| Proximity Measure | Accuracy Rate |
|---|---|
| MTDM (Euclidean based) | 80.61% |
| MTDM (Average based) | 75.15% |
| MTDM (City Block based) | 75.15% |
| MTDM (Chebychev based) | 72.73% |
| MTDM (Canberra based) | 72.12% |
| Simple Euclidean | 58.18% |

MTDM is a specialised trajectory distance measure for spatiotemporal (geospatial) applications which does not hold presumptions about the relationship between compared trajectories. This gives the approach its flexibility towards comparing all sorts of spatiotemporal trajectories (mainly involving geospatial types). MTDM also has multiple advantages including: capability to adapt the distance measurement to the relative scale of the problem (using the standardisation steps of the approach) and the capability to approximate trajectory distance in a comprehensive approach (utilising the multiple MTDM dimensions). MTDM can also compare different length trajectories and is not limited by the requirement to have equal number of points within the trajectories compared. Efficiency evaluation of MTDM showed the feasibility of applying the measure to various trajectory datasets as was tested in this research paper.

To test MTDM, a geospatial data analysis framework was utilised. This led to the effective visualization of geospatial data (moving-object trajectories) for the data mining experts as was required for the apriori-knowledge validation experiments.

For future research, improvement to the efficiency of the algorithms with spatial indexes can be tested. In addition, the robustness of the algorithm to noise can be tested as it is expected for the MTDM approach to behave well under "noisy" conditions. Testing alternative weighting of the dimensions of MTDM can also contribute to the theoretical foundation as this paper assumed equal weights of the dimensions. Integrating semantics from different spatiotemporal applications into the multi-dimensional distance measure can also be analysed in case-studies which can demonstrate the capabilities of MTDM in handling multiple spatiotemporal domains.

## ACKNOWLEDGMENT

This research study was funded by the research and development fund from Teradata Corporation, Egypt. Experiments were conducted within the laboratories of the German University in Cairo.

## REFERENCES

[1] GeoServer Software. (2012). Retrieved March 31, 2012, from GeoServer: http://www.geoserver.org

[2] *SQL Geospatial Types.* (2010). Retrieved March 1, 2012, from Teradata Website: www.info.teradata.com/edownload.cfm?itemid=102320050

[3] Trajectory Datasets. (2012). Retrieved March 31, 2012, from R-Tree Portal: http://www.rtreeportal.org

[4] Badawy, S. A., Elragal, A., & Gabr, M. (2008). Multivariate Similarity-Based Conformity Measure (MSCM): an Outlier Detection Measure for Data Mining Applications. AIA '08 Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications, (pp. 314-320).

[5] Buchin, K., Buchin, M., van Kreveld, M., & Luo, J. (2009). Finding Long and Similar Parts of Trajectories. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems GIS, (pp. 296-305).

[6] Chen, L., Özsu, M. T., & Oria, V. (2005). Robust and Fast Similarity Search for Moving Object Trajectories. Proceedings of the 2005 ACM SIGMOD international conference on Management of data, (pp. 491-502).

[7] Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., & Keogh, E. (2008). Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures. Proceedings of the VLDB Endowment, 1(2), 1542-1552.

[8] Domingo-Ferrer, J., Sramka, M., & Trujillo-Rasúa, R. (2010). Privacy-preserving Publication of Trajectories Using Microaggregation. SPRINGL '10 Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS , (pp. 26-33).

[9] Frentzos, E., Gratsias, K., & Theodoridis, Y. (2007). Index-based Most Similar Trajectory Search. IEEE 23rd International Conference on Data Engineering, ICDE 2007, (pp. 816 - 825).

[10] Frentzos, E., Gratsias, K., Pelekis, N., & Theodoridis, Y. (2005). Nearest Neighbor Search on Moving Object Trajectories. Lecture Notes in Computer Science, 3633/2005, 328-345.

[11] Funes, A., Ferri, C., Hernández-Orallo, J., & Ramírez-Quintana, M. J. (2009). An Instantiation of Hierarchical Distance-based Conceptual Clustering for Propositional Learning. Lecture Notes in Computer Science, 5476/2009, 637-646.

[12] Gordon, A. D. (1981). Monographs On Applied Probability And Statistics Classification. Cambridge: The University Press.

[13] İnan, A., & Saygın, Y. (2006). Privacy Preserving Spatio-Temporal Clustering on Horizontally Partitioned Data. *Lecture Notes in Computer Science, 4081/2006*, 459-468.

[14] Panagiotakis, C., Pelekis, N., & Kopanakis, I. (2009). Trajectory Voting and Classification Based on Spatiotemporal Similarity in Moving Object Databases. *Lecture Notes in Computer Science, 5772/2009*, 131-142.

[15] Pelekis, N., Kopanakis, l., Ntoutsi, I., Marketos, G., & Theodoridis, Y. (2007). Mining Trajectory Databases via a Suite of Distance Operators. IEEE 23rd International Conference on Data Engineering Workshop, (pp. 575 - 584 ).

[16] Rinzivill, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., & Andrienko, G. (2008). Visually–driven analysis of movement data by progressive clustering. Information Visualization, 7(3-4), 225-239.

[17] Vlachos, M., Gunopulos, D., & Kollios, G. (2002 ). Robust Similarity Measures for Mobile Object Trajectories. Proceedings of the 13th International Workshop on Database and Expert Systems Applications, (pp. 721 - 726).

[18] Vlachos, M., Kollios, G., & Gunopulos, D. (2002). Discovering similar multidimensional trajectories. Proceedings of the 18th International Conference on Data Engineering , (pp. 673 - 684).

[19] Yanagisawa, Y., Akahani, J.-i., & Satoh, T. (2003). Shape-Based Similarity Query for Trajectory of Mobile Objects. Lecture Notes in Computer Science, 2574/2003, 63-77.

 ISBN: 978-1-61208-251-6 

# Generation of Land Cover Maps Using High-Resolution Multispectral Aerial Cameras

Joachim Höhle
Aalborg University, Department of Planning
Aalborg, Denmark
jh@land.aau.dk

*Abstract*—**The generation of land cover maps has new advanced tools at its disposal. Besides new aerial cameras also new advanced processing tools are available. Elevations can be derived from images with high density and high positional accuracy. The combination of multispectral high-resolution imagery and high-density elevations for the automatic generation of land cover maps is discussed by means of a practical example. Imagery of a new aerial camera is used together with advanced software for generation of digital elevation models and for derivation of vegetation maps. These two products are the input for classification of land cover. A high degree of automation can be achieved. The obtained results of a practical example are checked with reference values derived from ortho-images in natural colour and from colour images using stereo-vision. An error matrix is applied in the evaluation of the results. The classification had an overall accuracy of 79%. Suggestions for the improvements in the applied methodology are made. The potential of land cover maps lies in updating of topographic databases, quality control of maps, studies of town development, and other geo-spatial domain applications. The automatic generation of land cover maps is also a trend in future mapping.**

***Keywords-land cover map; classification; assessment; multispectral camera; map revision.***

## I. INTRODUCTION

Automated classification of aerial images has to be able to distinguish between objects on the ground and objects above ground. It is difficult to distinguish between buildings and parking lots or trees and grass. Land cover maps can be generated with a higher semantic accuracy when information on elevations is available. The updating of topographic databases will profit from such an approach. Of special interest are the built-up areas where most of the changes occur and where the actuality of the map data is very much in demand. Usually, high-resolution images are used for the generation of topographic databases. Aerial cameras have improved considerably in the last 10 years. The new generation of aerial cameras has features which will make the generation of land cover maps easier and more reliable. The processing tools have improved too. Besides the professional (expensive) software tools nowadays open source software tools (freeware) are available. All the necessary tools have to be tested whether they fulfil the requirements of the practice. It is the goal of this contribution to use imagery of a new aerial camera and to produce land cover maps. In contrast to traditional approaches (where only intensity values of images are used) elevations will be used in addition. Three-dimensional point clouds are generated from high-resolution images with a very high density and positional accuracy. Furthermore, the multispectral images of the camera allow an automated classification of vegetation and other features. This approach for generating land cover maps has the potential for a fully automated processing. The land cover maps can be the base for updating of topographic databases and for several other geo-spatial domain applications.

The accuracy of classification has to be assessed. From the gained experiences recommendations for the generation of land cover maps and the revision of topographic databases will be given. The paper starts with some general remarks on the revision of map databases. The new generation of aerial cameras and processing tools is presented afterwards. A practical example of a land cover map generation is then carried out using up-to-date software tools including some own developments. The results are assessed by means of statistical methods. A conclusion is drawn and possible future work is outlined.

## II. GENERAL REMARKS ON THE REVISION OF MAP DATABASES

The revision of topographic databases requires detection of changes and errors. From the obtained results a decision has to be taken, in which way the renewal of the map database is to be carried out. It may be a completely new mapping, but mostly, an updating and correction of the existing information is carried out. This is done locally and in relatively small areas. It depends what type of data have to be revised. There are topographic databases in 2D and 3D. Sometimes only the important objects are updated, for example, buildings, roads, and trees. The process of updating should be an automatic process. This is at least the goal of today's map makers because stereo-plotting by operators is an expensive undertaking. Therefore, such work is often outsourced. In the past, a lot of efforts have been done to automate the detection of changes. Progress has been achieved using a semi-automated process. The automatically found changes will guide the operator to map new objects and to delete the obsolete ones. This may save some time for

 ISBN: 978-1-61208-251-6 

the operator, but the semi-automated process is not without costs, delays, and logistic problems. In order to automate also the mapping of the changes in 3D, the generation of a land cover map and of a digital elevation model (DEM) will be a first step. The automatic generation of land cover maps in 2D applies classification. The classification techniques use several attributes like spectral signature, texture, shape, etc. Also elevation has been taken as an attribute. If only elevations are used in classification, it is difficult to separate vegetated ground from parking lots and roads. The combined use of images **and** elevations is therefore a much better approach for the generation of land cover maps in 2D and 3D. The use of different data sources requires that all data are in the same reference system and that they have a good positional accuracy. Furthermore, all of the new data should be acquired simultaneously. Such conditions cannot always be fulfilled in practice. In addition, there are economic constraints that updating of topographic databases has to be carried out at low costs and in short time intervals.

In the past several studies have been carried out in the revision of maps, for example in [1]. In a project of the European Spatial Data Research (EuroSDR) different approaches were investigated [2]. Some research groups only used images, others used elevations only. The combination of images and elevations has also been suggested. Reference [3] used ortho-images of different times in order to derive areas of change. Reference [4] detected residential land use of buildings from lidar and aerial photographs through object-oriented classification. The event of a new generation of digital aerial cameras and of new processing software has changed the situation. It is therefore a goal of this paper to apply these new tools and to generate land cover maps automatically. The impact of such maps for the updating of topographic and other databases will also be discussed. Up-to-date data are the prerequisite in all geo-spatial domain applications. Land cover maps are often used for studies in town development. The changes in areas of buildings, traffic and vegetation over several years are studied in [5]. In that investigation the land cover maps are derived from vector maps and low-resolution satellite images. The applied classification of the images used intensity values only. The additional use of elevations of high density (as in this investigation) should result in more reliable results.

## III. NEW DIGITAL AERIAL CAMERAS

With the appearance of new advanced digital aerial cameras the generation of land cover maps has new tools at its disposal. There are different types and models of aerial cameras; only three of the new cameras will be discussed in the following: the Hexagon/Intergraph DMCII 250, the Microsoft UltraCam Eagle, and the Hexagon/Leica Geosystems RCD30 camera. Details of the cameras are described in recent publications [6] [7] [8]. All three are frame cameras, and they can produce images of high resolution and high geometric quality. They are designed for mapping tasks. The produced images have different features; the major ones are listed in Table I. There is a considerable difference in the format of the output image. The RCD30 is a medium-format camera; the other two cameras are considered as large-format cameras. A larger format requires less flying in order to cover an area to be mapped assuming that the images of all cameras have the same ground sampling distance (GSD). The necessary length of the flight is an economic factor, and large-format cameras have an advantage in projects covering large areas. All three cameras can produce black&white, colour and false-colour images simultaneously. Newly designed lenses match the resolution of the sensor(s) and enable high image quality. The cameras are calibrated, and the obtained calibration data are used to correct the images in geometry and radiometry. Additional sensors for position (GNSS) and attitude (IMU) can be supplemented and will support accurate georeferencing of the imagery.

TABLE I. FEATURES OF THREE NEW DIGITAL AERIAL CAMERAS.

| Features | | UC Eagle | DMCII 250 | RCD30 |
|---|---|---|---|---|
| pixel size | [µm] | 5.2 | 5.6 | 6.0 |
| focal length | [mm] | 80 | 112 | 50, 80 |
| image size(in flight direction) | [pel] | 13 080 | 14 656 | 6708 |
| | [mm] | 68.0 | 82.1 | 40.2 |
| image size(across flight direction) | [pel] | 20010 | 17216 | 8956 |
| | [mm] | 104.1 | 96.4 | 53.7 |
| number of pixels per image | [MP] | 262 | 252 | 60 |

Because the RCD30 images will be used in the following practical example some more details have to be mentioned. The colour images are produced from one CCD with Bayer filters. The infra-red band is imaged by a second CCD of the same high resolution (pixel size=6 µm x 6 µm). Image motion is compensated mechanically in two axes.

The output images are corrected for distortion, light-fall off of the lens and non-uniformity for dark signals. Two different lenses can be used without the need of re-calibration. In order to obtain images with GSD=0.05 m they have to be taken from 417 m above ground with a 50 mm lens. One frame will cover 0.15 $km^2$ on the ground. A gyro-stabilized mount can be used which will prevent big tilts of the imagery.

The potential of the mentioned new cameras with respect to automatically derived elevations has been investigated in [9].

## IV. PROCESSING TOOLS

The generation of land cover maps from images requires various software tools. They reach from general image processing to dedicated software for photogrammetry, remotes sensing and GIS. One of the problems to overcome is the big amount of data. The work with small units may be necessary. One 64-bit computer with 8 Gb RAM (as it was at disposal in this investigation) is not an optimal processor. More computer power is necessary to solve the task for large areas. The many different tools are expensive if they have to be acquired from software vendors. Some software was available as freeware and open source. Such tools were preferred for this task. Some programming has been carried out using R-language, which is also available as open source.

 ISBN: 978-1-61208-251-6

## V. PRACTICAL EXAMPLE

A practical example has been carried out in order to evaluate the proposed approach. A residential area in Switzerland with buildings, roads, paved paths, parking lots as well as trees, hedges and grassland was photographed. The size of the area is about one hectare. Errors and changes in topographic objects should be discovered. Of special interest are the errors and changes in the man-made objects.

### A. Outline of the methodology

The land cover map to be produced should comprise the following classes: 'buildings', 'trees&hedges', 'grass', 'roads&parking lots'. These objects have certain attributes which will be used in the classification process. The used attributes are elevation and vegetation, which can automatically be derived from images. False colour images allow a separation of the image content in vegetated and non-vegetated areas using the Normalized Density Vegetation Index (NDVI). True colour images are used to derive a very dense 3D point cloud by means of matching corresponding image parts. The very dense point cloud has to be transformed into a Digital Surface Model (DSM). Filtering and editing will generate a Digital Terrain Model (DTM). The difference between the DSM and the DTM is the normalized Digital Surface Model (nDSM). NDVI data and nDSM data have to be in the same reference system. The DTM is also used for the production of a false colour ortho-image which is further processed to a NDVI map containing two classes (vegetation, non-vegetation). NDVI and nDSM information is then used to produce a land cover map. Fig. 1 depicts the steps in the production of the land cover map. Details of the classification process are also presented in a flowchart (cf. Fig. 2). Separation into ground and above ground or vegetated and non-vegetated areas is realized by specifying thresholds. At the position of the DSM points the NDVI attributes (vegetated or non-vegetated) are assigned. The classification of all DSM points with their attributes (ground, above ground, vegetated, non-vegetated) is then carried out and the result is plotted using different colours for the selected classes ('buildings','trees&hedges','grass','roads&parking lots').

### B. Data

The used RCD30 imagery includes four bands (red, green, blue, near infrared) from which colour and false-colour images can be composed. Each image has about 6.4 Megabyte (Mb). The images were taken with a GSD of about 5cm. The images are geo-referenced and data of the geometric calibration (camera constant, pixel size) have been provided. As reference data a colour ortho-image has been generated. Furthermore, a map of building footprints was produced by digitizing from a stereo-pair of colour images.

### C. Applied tools

For the generation of a 3D point cloud the Match-T program, version 5.4, of the Trimble/Inpho Company could be used [10]. Filtering of the data has been carried out automatically using filters of the professional program "DTMaster", version 5.4. Some manual editing was necessary and has been carried out by the same program. The orthoimages in false colours have been produced by Inpho's "OrthoMaster" program. The difference between the two elevation models (nDSM) has been calculated by a program produced in C-language. The generation of the land cover map was carried out by an R-program. Input was the nDSM and the NDVI map. The NDVI map was generated by the open source software LEOWorks, version 4.0 [11]. The produced land cover map was graphically improved by means of the Quantum GIS program, version 1.7.4, which is an open source Geographic Information System [12]. The assessment of the land cover map used the DTMaster program which enabled dynamic positioning to derived sample data.

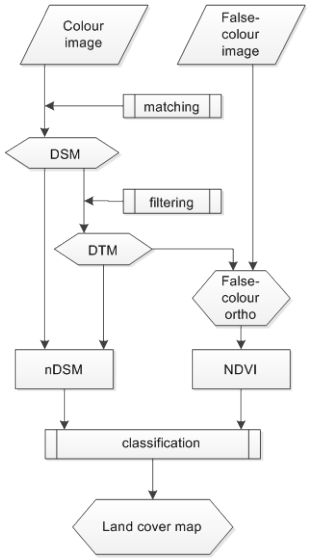



Figure 1. Steps in the production of a land cover map.

### D. Results

The calculation of the 3D point cloud revealed a theoretical height accuracy of $\sigma_Z$ = 0.16 m. The absolute accuracy of the DSM was determined by means of a few check points and resulted in $RMSE_Z$ = 0.22 m. The obtained nDSM had a relatively high density (up to 16 points/$m^2$). Its accuracy is estimated with $RMSE_Z \approx$ 0.3m. The NDVI map had two levels which were separated by a threshold of NDVI > 0.1. The threshold for separating low and tall vegetation or non-vegetated objects was selected with 1.0 m. The result of the classification using these two inputs is depicted in Fig. 3. The four classes are plotted by different colours and symbols and can be well separated from each other. This land cover map is georeferenced and can therefore be overlaid by a map of the building footprints (cf. right part of Fig. 4). The land cover map is based on the numerical output of the classification. Besides the spatial coordinates (Easting, Northing, elevation ($Z_{DSM}$)), the difference in

 ISBN: 978-1-61208-251-6 

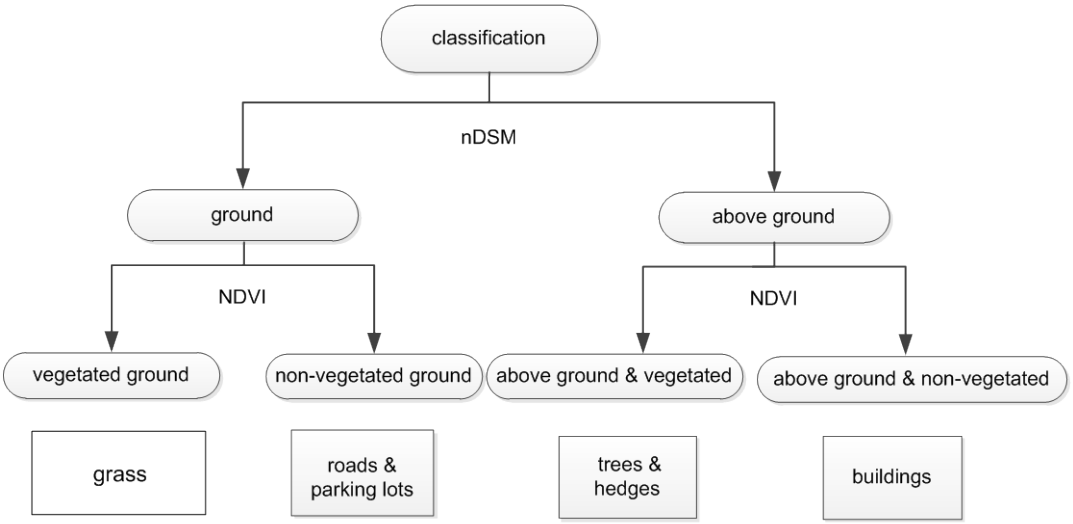



Figure 2. Classification flowchart.

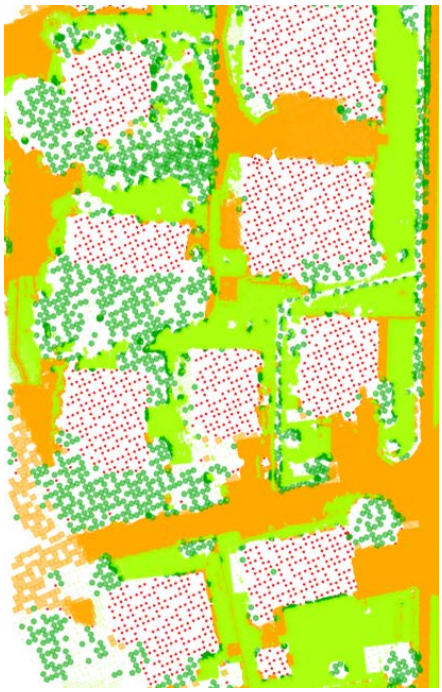


Figure 3. Result of classification. Legend: Red dots (‘buildings’), brown dots (‘roads&parking lots’), bright green dots (‘grass’), dark green dots (‘trees&hedges’).

elevation (dZ) between the DSM and DTM and the type of the point (class) are contained in the listing (cf. Table II). The list represents a “classified point cloud” and can be used for other purposes than mapping as well.

### *E. Checking of the classification accuracy*

The classification can be checked visually. Fig. 4 depicts the input data together with the result. The overlay of

TABLE II. EXAMPLE OF A “CLASSIFIED POINT CLOUD”.

| Easting | Northing | Z | dZ | class |
|---|---|---|---|---|
| 537129.2 | 5228938.6 | 486.5 | 0.2 | “grass" |
| 537129.2 | 5228938.7 | 488.5 | 2.3 | "trees&hedges" |
| 537144.5 | 5228987.4 | 486.4 | 0.0 | “roads&parking lots" |
| 537128.0 | 5228938.3 | 490.8 | 4.2 | "building" |

building footprints with the land cover map shows pretty good agreement. The buildings and the trees are well separated. The shaded DSM and the land cover map display even small changes in elevations. For example, hedges can clearly be recognized in the land cover map.

Errors in the separation between grass land and paths can be noticed in the shadows of houses. A few differences in the buildings can also be recognized. The manual mapping of the buildings used some generalizations with the result that balconies and extensions are missing. This is evident in the overlay of building outlines onto the land cover map.

The numerical assessment of the result can be done by means of reference values. These are derived by manual interpretation of a reference. For each class a number of DSM-points are randomly selected and compared with the reference. The accuracy of the classification is derived by applying an error matrix as suggested in [13]. The achieved classification accuracy is derived overall and for each class. It is distinguished between producer’s and user’s accuracy.

As reference, the colour ortho-image, which is based on the DTM, is chosen first. The classification has been verified at 91 randomly selected samples per class, altogether at 364 positions respectively. The sum of reference values in each class is listed as rows of the error matrix (cf. Table III). For example, at 91 sample positions of the class ‘buildings’ 43 were identified as ‘buildings’, 31 as ‘roads & parking lots’, 5 as ‘trees&hedges’ and 12 as ‘grass’ by means of a manual determination of the reference.

 ISBN: 978-1-61208-251-6 

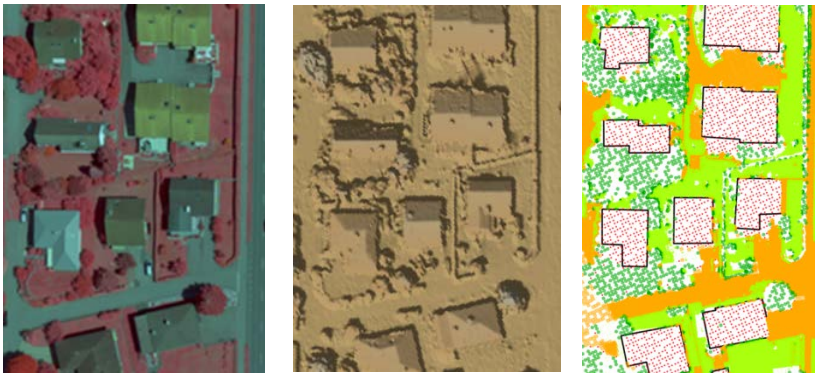


Figure 4. False colour ortho-image, shaded DSM and land cover map with overlay of building outlines.

TABLE III. ERROR MATRIX OF THE DERIVED LAND COVER MAP BY EVALUATION OF TRUE COLOUR ORTHO IMAGES.

| | 1 | 2 | 3 | 4 | Σ |
|---|---|---|---|---|---|
| 1 buildings | 43 | 31 | 5 | 12 | 91 |
| 2 roads&parking lots | 0 | 78 | 1 | 12 | 91 |
| 3 trees&hedges | 1 | 6 | 50 | 34 | 91 |
| 4 grass | 0 | 6 | 8 | 77 | 91 |
| Σ | 44 | 121 | 64 | 135 | 364 |

The overall accuracy of the classification is 68%. Of interest are the results for each class. The producer's accuracy for each class is 98% ('buildings'), 64% ('roads & parking lots'), 78% ('trees&hedges'), and 57% ('grass'). The user's accuracy of 'buildings' is 47%, of 'roads&parking lots' 86%, of 'trees&hedges' 55%, and of 'grass' 85%.

More accurate reference values can be obtained by means of stereovision using true colour images. The results are given in Table IV. The classification has been calculated again. The overall accuracy is 79%. The producer's accuracy for each class is 94% ('buildings'), 85% ('roads&parking lots'), 60% ('trees&hedges'), and 63% ('grass'). The user's accuracy of 'buildings' is 70%, of 'roads&parking lots' 92%, of 'trees&hedges' 59%, and of 'grass' 94%.

For the evaluation of the result the Kappa analysis is applied. Furthermore, the variance of Kappa (K) and the significance of the error matrix (Z statistic) are calculated by means of formulae published in [13]. The results are presented in Table V.

TABLE IV. ERROR MATRIX OF THE DERIVED LAND COVER MAP BY EVALUATION UNDER STEREO VISION OF TRUE COLOUR IMAGES.

| | 1 | 2 | 3 | 4 | Σ |
|---|---|---|---|---|---|
| 1 buildings | 64 | 12 | 15 | 0 | 91 |
| 2 roads&parking lots | 1 | 84 | 3 | 3 | 91 |
| 3 trees&hedges | 3 | 1 | 27 | 15 | 46 |
| 4 grass | 0 | 2 | 0 | 30 | 32 |
| Σ | 68 | 99 | 45 | 48 | 260 |

TABLE V. ERROR MATRIX AND RESULT OF KAPPA ANALYSIS.

| Error matrix | $\hat{K}$ | Variance | Z statistic |
|---|---|---|---|
| analysis #1 (ortho) | 0.58 | 0.00100 | 18.2 |
| analysis #2 (stereo) | 0.71 | 0.00116 | 20.7 |

## VI. DISCUSSION

The achieved overall accuracy derived from the 364 or 260 samples is 68% and 79% respectively. The user's accuracy is more important than the accuracy of the producer and will therefore be discussed further. The classification of buildings with 47% is poor when the evaluation is done by means of the true colour ortho-image. Much better is the result for 'roads&parking lots' and 'grass' (86% and 85% respectively). The class 'trees&hedges' is correct by 55%. When evaluation is done by means of colour images under stereovision (which requires the Z-coordinate of the sampled data) 'buildings' are classified with 70% which is a much better result. The classification of 'roads&parking lots' is achieved with even 92%. According to [14] the Kappa values (0.58 and 0.71) represent a moderate agreement between the classification and the reference data. The derived variances are small; the Z-statistic–value is far above 1.96 which represents a 95% confidence level. This means that the applied classification is significantly better than random. The comparison of the analysis #1 with analysis #2 shows about the same result for the Z-statistics (18.2 and 20.7). A statistical test indicated that the difference between the two Kappa values is significant. The results may be further improved by applying a bigger threshold than 1m in elevation or by additional steps in the elevations in order to separate more classes, for example shrubs, hedges, walls, etc. Also corrections for shadows and displacements of elevated objects in standard ortho-images could be considered. That means that true ortho-images instead of standard ortho-images could be used. Economic considerations will decide this question. The use of true ortho-images needs digital building models (DBMs), which are not generally available. The production of accurate DBMs is pretty expensive. The applied evaluation by means of stereovision is very useful.

 ISBN: 978-1-61208-251-6 

## VII. EVALUATION

In this paper elevation data of high density have been derived from aerial images. The applied high-resolution multispectral images also enabled an automatic classification of vegetation. By means of a combined use of elevation and vegetation data a land cover map could be produced with a high degree of automation. Such a map is the graphical output of a "classified point cloud" of high density. In the example the attributes of the point cloud were the spatial coordinates (Easting, Northing, and elevation), the normalized height and four classes (grass, roads and parking lots, trees and hedges, and buildings). The "classified point cloud" may have more classes when an extended classification algorithm will be used. For example, steps in the normalized height can differentiate various types of vegetation. In addition, water areas can automatically be extracted from the near-infrared channel of the multispectral imagery.

Decisive for good results of land cover maps is the quality of the applied imagery. A measure for the image quality is the so-called Point Spread Function (PSF). It is derived from edges within the image. The width of the PSF is considered as 'scale factor of the effective pixel size' [15]. This factor has been derived for one of the applied images with k=1.27 (average between centre and edge of the image). Such a value is close to values found for large-format aerial cameras [15]. In addition, the so-called Photo Response Non Uniformity should be checked. In tests of the applied imagery it was found that the intensity values were nearly the same in the centre and the edge of the image. This means that the classification of vegetation using NDVI has had prerequisites for good results. This could be confirmed by checking the vegetation map where only 6.4% errors were found. All of these investigations revealed good results for the applied imagery. The use of a medium-format digital camera (which is considerably less expensive than a large-format camera) is also a novel approach for the generation of land cover maps.

Digital Terrain Models, Digital Surface Models, as well as original 3 D points clouds may already exist for large areas. The generation of land cover maps may use such elevation data of existing databases. However, a quality control should be carried out whether the data are "fit for purpose". High density, high positional accuracy, and completeness of the elevation data are prerequisites of the presented approach. Many other data in existing databases may be useful for the generation of land cover maps. The extraction of various data from existing databases and combining with new imagery is one of the big challenges in GIS technologies.

## VIII. CONCLUSION AND FUTURE WORK

Land cover maps are a mapping product which has many applications. They can be used for the updating of topographic databases and for the quality control of topographic databases. Other applications are establishing of tree cadastres in municipalities and general studies in town development. The new tools like advanced aerial cameras and processing software packages make the automatic generation of land cover maps easier and more accurate. Especially, the use of elevations yields better results in the classification. The quality control by means of stereo-observations has to be used in order to receive reliable reference values. The described approach for the generation of land cover maps can be further improved by using information of existing databases. The number of classes can then be increased. Such an approach should be applied for large areas and comprise different types of urban areas.

## ACKNOWLEDGEMENT

The author thanks Hexagon/Leica Geosystems for providing image data. Trimble/Inpho and K. Jacobsen, Leibniz University Hannover, are thanked for loaning of software. M. Höhle supported the author in programming. B. Nørskov is thanked for improving the English language.

## REFERENCES

[1] B. P. Olsen, "Maintenance of digital topographical map databases—change detection," PhD thesis, Technical University of Denmark, Lyngby, 2004, 122p.

[2] N. Champion, "Detection of unregistered buildings for updating 2D databases," EuroSDR Official Publication, no. 56, 2009, pp. 7-54.

[3] A.A. Nielsen and B. P. Olsen, "Change detection for semi-automatic map database updating," National Survey and Cadastre, Denmark, technical report series number 09, 2010, 24 p.

[4] X. Meng, N., Currit, L. Wang, and X. Yang, "Detect residential buildings from lidar and aerial photographs through object-oriented land-use classification", Photogrammetric Engineering and Remote Sensing, vol. 78, no. 1, 2012, pp. 35-44.

[5] L. Halounova, K. Veprek, and M. Rehak, "Geographic information systems models of 40-year spatial development of towns in the Czech Republic", Proc. GEOProcessing, 2011, pp. 75-80.

[6] K. Jacobsen and K. Neumann, "Property of the large-format digital aerail camera DMC II", Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XXXIX-B1, 2012, pp. 21-25.

[7] M. Gruber, M. Ponticelli, R. Ladstädter, and A. Wiechert, "UltraCam Eagle, details and insight", Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XXXIX-B1, 2012, pp.15-19.

[8] R. Wagner, "The Leica RCD30 medium-format camera: imaging revolution," Proc. Photogrammetric Week '11, Wichmann Verlag, 2011, pp. 89-95.

[9] J. Höhle, "On the potential of new digital aerial cameras for DEM generation,"The Photogrammetric Journal of Finland, vol. 22, no. 2, 2011, pp. 27-36.

[10] T. Heuchel, A. Köstli, C. Lemaire, and D. Wild, "Towards a next level of quality DSM/DTM extraction with Match-T," Proc. Photogrammetric Week '11, 2011, Wichmann Verlag, pp. 197-202.

[11] ASRC, "LEOWorks, version 4.0," 2011. http://leoworks.asrc.ro (accessed 23.07.2012)

[12] QGIS, "User Guide of Quantum GIS program, version 1.7.0," 207 p., 2011.
http://download.osgeo.org/qgis/doc/manual/qgis1.7.0_user_guide_en.pdf (accessed 21.07.2012)

[13] R. G. Congalton and K. Green, "Assessing the accuracy of remotely sensed data," CRC Press, 183p., 2009, ISBN 978-1-4200-5512-2.

[14] J.Landis and G. Koch, "The measurement of observer agreement for categorical data", Biometrics, vol. 33, 1977, pp. 159-174.

[15] K. Jacobsen, "Tells the number of pixels the truth?-Effective resolution of largesize digital frame cameras", Proc. ASPRS Annual Conference, Portland, Oregon, 2008, 7p.

 ISBN: 978-1-61208-251-6 

# On Improving Geotag Quality in Photo Collections

Yuri Almeida Lacerda[1,2], Hugo Feitosa de Figueirêdo[1,3], João Paulo Ribeiro da Silva[1], Daniel Farias Batista Leite[1], Cláudio de Souza Baptista[1]
[1]University of Campina Grande
Campina Grande,Paraíba, Brazil
[2] Federal Institute of Education, Science and Technology of Ceará - Campus Crato - Crato, Ceará, Brazil
[3] Federal Institute of Education, Science and Technology of Paraíba – Campus Monteiro - Monteiro, Paraíba, Brazil
yurilacerda@ifce.edu.br, hugo.figueiredo@ifpb.edu.br, joao.silva@ccc.ufcg.edu.br, daniel.leite@ccc.ufcg.edu.br,baptista@dsc.ufcg.edu.br

Anselmo Cardoso de Paiva[4]
[4] Federal University of Maranhão
Applied Computing Group NCA
São Luís, Maranhão, Brazil.
paiva@deinf.ufma.br

***Abstract*—Digital cameras equipped with GPS receivers allow storing geographic location into the photograph metadata. Geographic location constitutes a very important information to be used in systems that retrieve and organize photographs. However, cameras equipped with GPS may store either invalid or null geographic locations. This is usually due to a delay in obtaining the GPS signal. In this paper, we propose a new method for detection of inconsistencies of geographic locations in photograph metadata and for propagation of the geographic location annotation for these photographs. Besides, we also propose the incorporation of these methods in a Web-based system for correction and annotation of geotags. Another contribution of this work is the presentation of a case study to validate the system. The results prove that the proposed techniques increase precision (through the detection of inconsistencies) and recall (through the propagation of the geographic locations) in the retrieval of georeferenced photographs.**

***Keywords* - Geotag; Photo Metadata; GeotagPropagation; Metadata Inconsistency Detection.**

## I. INTRODUCTION

The technological advances in the last years have enabled a wide use of electronic devices, such as: digital cameras, smartphones, and tablets. This popularization caused an exponential increase of the amount of multimedia files produced by people, such as videos, photographs and audio. This phenomenon can be easily verified in social networks, blogs and internet sites. The large number of multimedia files that has been generated by people has jeopardized information management. For example, imagine an user organizing manually a collection with thousands of photographs. Even manually organizing hundreds of photographs taken during a vacation trip is very time consuming and tedious task.

Several approaches have been proposed for automatic organization of photographs, with the objective of reducing the user's manual efforts, such as PhotoGeo [1], Naamanet al. [2], Cooper et al. [3] and Tsay et al. [4]. Such systems usually make use of the photo metadata to help in the organization process. Some examples of these metadata include date, time, geographic location of the camera at the moment of capture, camera manufacturer and model, tags and descriptive data.

Some studies argue that the place where the photograph was taken is one of the first things people remember when they want to retrieve that photograph [2]. This means that the geographic location of the camera at the time the picture was taken is very important for the process of photograph organization.

The integration of GPS chips into smartphones and digital cameras have allowed the storage of geographic location in the metadata of photographs automatically. Nevertheless, there can be some problems related to this information acquisition such as data imprecision, invalid data and indoor difficulties.

The low power of the GPS chips supplied with those devices, and the poor quality of the GPS signal in many places may generate imprecise geographic location data. Thus, the photographs may end up being indicated in places far away from the real point where the picture was taken.

Another problem is the fact that GPS receivers do not work well indoors, possibly generating invalid or imprecise data in this situation. Some smartphones uses the A-GPS [5] system to minimize this problem.

In other situations, the georeferencing is either absent or taken erroneously. For example, suppose that a given person with a camera equipped with GPS has captured some photographs. We know that it is necessary some instants until the chip receives the GPS signal. While the signal is not received, the camera will not make the georeference of the photographs, or it may use the last geographic information captured by the GPS, possibly generating incorrect information. However, if the user remains with the camera on and the GPS function activated, within a few instants the photographs will be taken with correct geographic information.

In this work, we propose an automatic and semiautomatic photograph georeferencing system based on the detection of inconsistencies of geographic location data and correction propagation, with the objective of augmenting the georeference recall and precision in a photograph collection. The proposed

 ISBN: 978-1-61208-251-6 

system uses temporal segmentation and the geographic location of some georeferenced photographs to detect inconsistencies in the locations. The system also suggests new annotations for the photographs with mistakes or without annotation.

The remaining of the paper is structured as follows. In Section II, we highlight some studies related to the subject approached in this paper. Next, Section III focuses on the prototype architecture. Section IV addresses the solutions proposed in this paper, for detection of inconsistencies and suggestion of geographic location annotation. Section V presents the evaluation of the proposed solutions through experiments and analysis of the results. Finally, in Section VI, we present the conclusion and discuss further work to be undertaken.

## II. Related Work

In this section, we discuss related work. Initially, we present studies dealing with the use of tags. Next, we focus on the use of content to make the images georeferenced and, finally, we present the studies on propagation of geotags.

Lee et al. [6] highlight the existence of a strong correlation between purely textual tags and the geographic location of the photographs in social networks. So, in the proposed approach, a computation of the similarity between tags and geographic location of the photographs was used to determine the relationship between the tags and geotags.

Hays and Efros [7] propose an algorithm called "im2gps", that estimates the geographic location of a photograph based on the geographic location of photographs with higher visual similarity. For such, they used a database containing more than 6 million georeferenced photographs.

Hollenstein and Purves[8] carried out a study demonstrating that the geotag must be according to the way people describe a place, that is, instead of georeference through latitude and longitude, tags like "Eiffel Tower", for example, should be added.

Many studies deal with the propagation of tags based on geographic location [9] and infer the geographic location based on the image content and tags [10].

Vandormael and Courdec[11] used the communication between devices to check the coherence of the geographic location of mobile devices at the moment the photographs are captured.

Ivanov et al. [9] proposed the propagation of geotags based on the combination of detection of duplicated objects and the user's confidence modeling. The idea is the propagation of geotags using other geotagged photographs.

CrEve [12] is a collaborative event annotation framework that uses photograph content found in social media sites. One of the addressed issues is inconsistency of photo metadata. However, the user must create an event using social media and this framework does not focus on personal photographs.

In this work, we propose a new automatic and semiautomatic scheme for georeferencing photographs that uses other already georeferenced photographs from the same personal collection. Furthermore, we propose a scheme for detection of inconsistencies in the photograph georeferencing that uses the spatiotemporal dimension of the collection. To the best of our knowledge, there is no work in the literature with a similar approach.

## III. System Architecture

The proposed approach was integrated to PhotoGeo [1]. PhotoGeo is a digital multimedia library specialized in georeferenced photographs. It has a multilayer architecture and was developed in compliance to the MVC (Model-View-Controller) design pattern.

Figure 1 presents the PhotoGeo architecture, highlighting the data, business logic and view layers. In this work, we added the Photograph Georeferencing module, responsible for integrating the detection of inconsistencies and georeferencing propagation. This module is described in detail the next section.

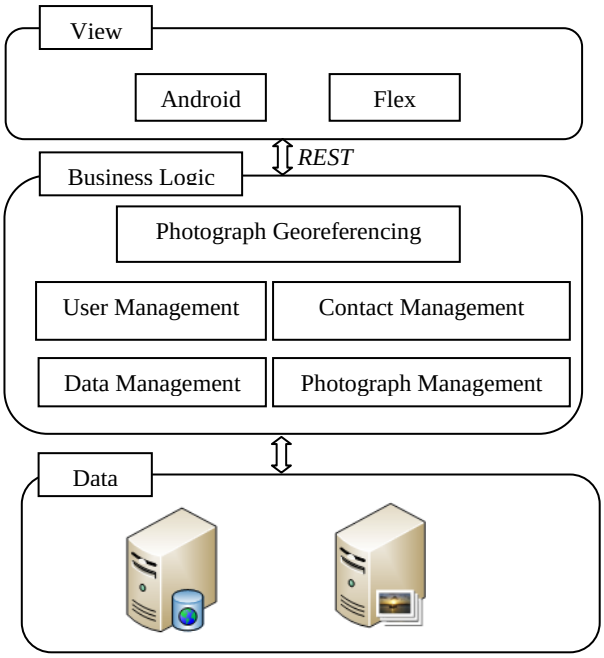



Figure 1. Prototype Architecture.

The data layer comprises two databases: an object-relational database (PostgreSQL) with spatial support (PostGIS); and the user photograph collections.

The business logic layer comprises the following main modules: Data Management, User Management, Contact Management, Photograph Management and Photograph Georeferencing. These modules will be detailed next.

Data Management is responsible for accessing and mapping of data in objects. In this module, JPA (Java Persistence API) is used to make the data persistence.

User Management and Contact Management modules are responsible for managing of system users and contacts, respectively.

The Photograph Management module is in charge of inserting, removing and retrieving photographs, besides

 ISBN: 978-1-61208-251-6 

extracting metadata. Moreover, in this module the access permissions for the photographs are checked.

The communication between the view layer and the business logic layer is made through REST (Representational State Transfer). Figure 2 shows the prototype interface used to present inconsistencies, and suggestions for the correction and georeferencing of photographs.

## IV. Detection and Correction of Inconsistencies

In this section, we present the solution proposed for propagation and detection of inconsistencies in geotags.

### A. Geotags Propagation

In this subsection, we present the solution proposed for geotag propagation from georeferenced photographs to non georeferenced ones, in the same personal collection. First, we perform a temporal segmentation on the set of photographs, using the $t_{\max}$ segmentation time input parameter, in minutes.

Assuming that F is a set of $n$ photographs, the temporal segmentation is responsible for separating the photographs into $k$ non intersecting clusters $g$, in such a way that:

$$\left(\bigcup_{j=1}^{k} g_j\right) = F \quad (1)$$

This segmentation produces clusters whose photographs have a maximum temporal difference of $t_{max}$ minutes between two temporal consecutive photographs. That is, considering that $f_i$, and $f_{i+1}$ are two consecutive photographs in the cluster $g_k$, and that $t_i$ and $t_{i+1}$ are their timestamps, then $t_{i+1} - t_i \leq t_{max}$. The photograph collection will be segmented based on Equation (2) and each photograph will belong to exactly one cluster.

$$\begin{array}{c} f_1 \in g_1 \\ f_{i+1} \in g_s \text{if } (t_{i+1} - t_i) \leq t_{max} \\ f_{i+1} \in g_{s+1} \text{if } (t_{i+1} - t_i) > t_{max} \end{array} \quad (2)$$

For every cluster $g_k$ and considering $fg_i$ as the geographic location of the photograph $f_i$, the iteration is made in order to find the non georeferenced photographs ($fg_i = \emptyset$). For each photograph with no geotag, $f_r$, we search for a photograph that is temporally closer, $f_s$, inside the same cluster, and that is georeferenced ($fg_s \neq \emptyset$). In these cases, photographs with GPS state in interoperability mode are not considered. This mode indicates that the geotag of the photograph may be imprecise. The geotag will be propagated from the photograph $f_s$ to $f_r$. The propagation can be automatic, without user interaction, or semiautomatic, when the user may accept or reject the suggestion.

When the propagation occurs, the same procedure is done, recursively, for other non georeferenced photographs. So, the new geotag of $f_r$, which was propagated from $f_s$, may be propagated to other photographs.

### B. Geotag Inconsistence Detection

In this subsection, we present the proposed algorithm for detection of geotag inconsistencies in photographs. This algorithm iterates on a subset of photographs, $F$, in a personal collection of a certain user, locating the georeferenced photographs. For each georeferenced photograph, $f_i$, we retrieve the photograph $f_j$, that has the smallest timestamp difference with respect to $f_i$. Right after that, the *maximum tolerable spatial distance* (*mtsd*) between the photographs is computed, through Equation (3). This computation is made by computing the difference from $f_i$ timestamp ($t_i$) to $f_i$ timestamp ($t_j$), and multiplying it by the *mean shift speed* (*mss*).

$$mtsd = |t_i - t_j| \times mss \quad (3) \quad (3)$$

The *mss* is the speed of the camera. However, there can be adverse circumstances, for example, in the case of a photograph being captured from inside an airplane which travels at 900 km/h. In this case, the photographs will have a considerable spatial distance, due to the shift speed. The same problem could happen when pictures are taken from inside a train, car or any other kind of vehicle which travels at high speed. For these situations, it is possible to capture speed information from the GPS chip. The *mss* must be supplied as one of the input parameters of the algorithm.

Right after that, we compute the *spatial distance* (*d*) between the photographs $f_i$ and $f_j$, using their geotags. In the case $d$ is greater than $mtsd$, the algorithm points out that there is a georeference inconsistence between the photographs, because it is unlikely that, moving from $ft_j$, with a mean speed *mss*, the user reaches the location of $fg_i$.

$$\begin{array}{c} d = distance(f_i, f_j); \\ \text{if} \begin{cases} d > mtsd, inconsistence \\ d \leq mtsd, \ consistence \end{cases} \end{array} \quad (4)$$

## I. Evaluation

In this section, we present the experiments carried out to validate the solutions proposed in this work for geotag propagation and inconsistency detection, respectively. Besides, we also present the methodologies used to perform the experiments.

### A. Geotag Propagation

In order to validate the geotag propagation scheme, we performed experiments to compare the geotags propagated by the algorithms, both in automatic and semiautomatic mode, with the real location of the non georeferenced photographs.

 ISBN: 978-1-61208-251-6

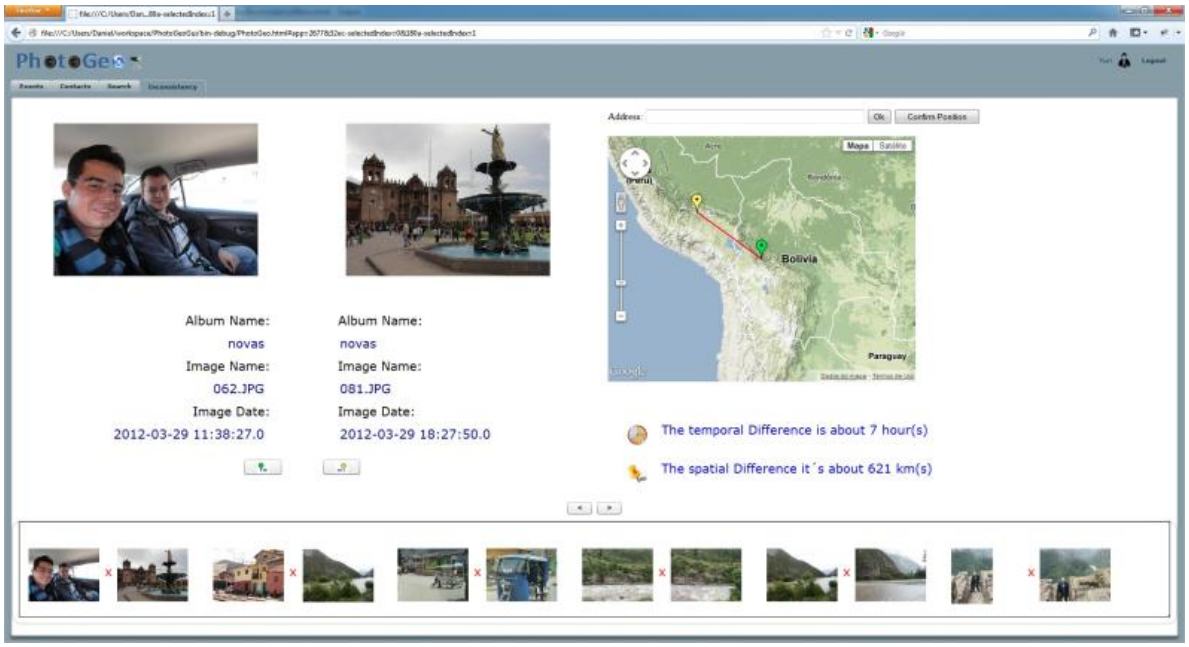



Figure 2. Photograph georeferencing interface.

To carry out the experiments, we used the precision and recall metrics. Precision is defined as the ratio between the correct propagation and all the photographs with geotag propagated. The propagation is considered correct when it is within a *maximum tolerable distance* (*mtd*), in meters, from the real geographic location of the photograph (informed by the user). On the other hand, recall is computed as the ratio between the correct propagation to the all non-georeferenced photographs in the collection.

To perform the experiments, we used a collection containing 4,153 photographs, from which 503 (12.11%) are not georeferenced. In order to automate the experiments, we obtained manually from the users the correct location of each photograph using a map.

Figure 3 presents the results for the computation of precision in semiautomatic mode, varying the parameters *mtd* from 10 to 300, with a step of 10. Each line in that graphic represents an experiment for a different $t_{max}$. The parameter $t_{max}$ varied from 5 to 60 minutes, at every five minutes. Figure 4 presents precision for the automatic mode, with the same variations of the parameters *mtd* and $t_{max}$.

We notice that precision increases when the segmentation time $t_{max}$ falls, because the longest the temporal distance between the photographs in the cluster is, more distant will be the propagated geotag, that is, it becomes more imprecise. With respect to the *mtd* parameter, precision is directly proportional, because the longest the tolerable distance is, more propagations will be considered to be correct.

In both cases, the best value of the $t_{max}$ parameter was five minutes. So, we chose the value to analyze the best precision for both modes. Thus, in semiautomatic mode, we achieved precisions varying from 94.89% to 97.08%. On the other hand, in automatic mode, we achieved precisions between 92.05% and 96.97%. We notice, then, in both cases, high precision was achieved, but semiautomatic mode gave better results.

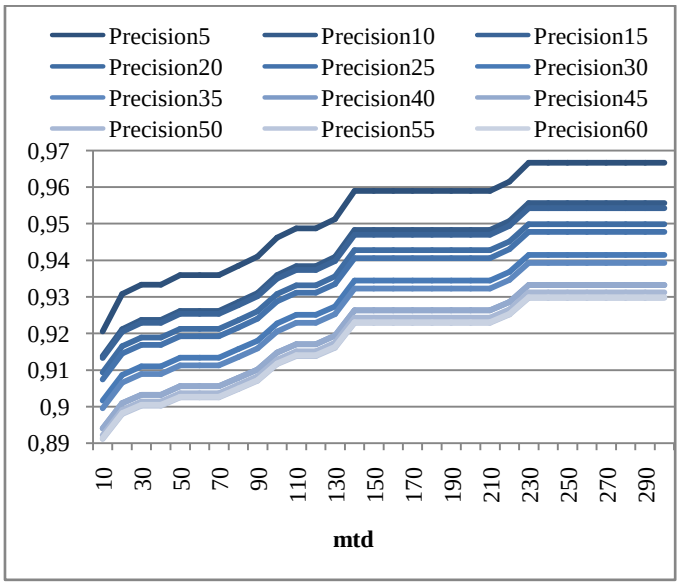



Figure 3 - Precision for the propagation experiment in semiautomatic mode.

For recall, using 5 minutes for $t_{max}$, we did not achieve good results (71.3% - 74.9% in semiautomatic mode, and 71.1%-72.7% in automatic mode). In Figure 5 we illustrate the result for recall in semiautomatic mode and in Figure 6 for automatic mode. Each line in the graphics represents a variation of the $t_{max}$ parameter. It can be noticed that recall is inversely proportional to $t_{max}$. This happens because the longer is the temporal distance between photographs in a cluster, the higher will be the number of geotag propagations in

 ISBN: 978-1-61208-251-6 

this cluster. So, the best value for $t_{max}$ to maximize recall was 60 minutes.

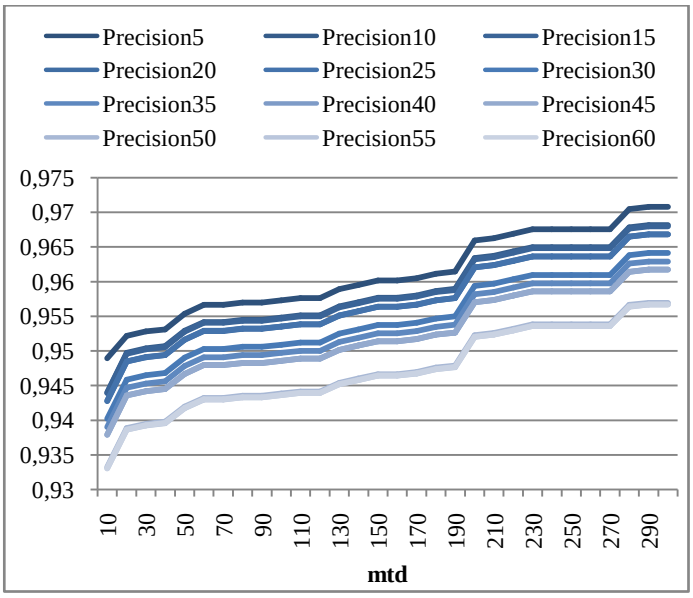



Figure 4 - Precision for the propagation experiment in automatic mode.

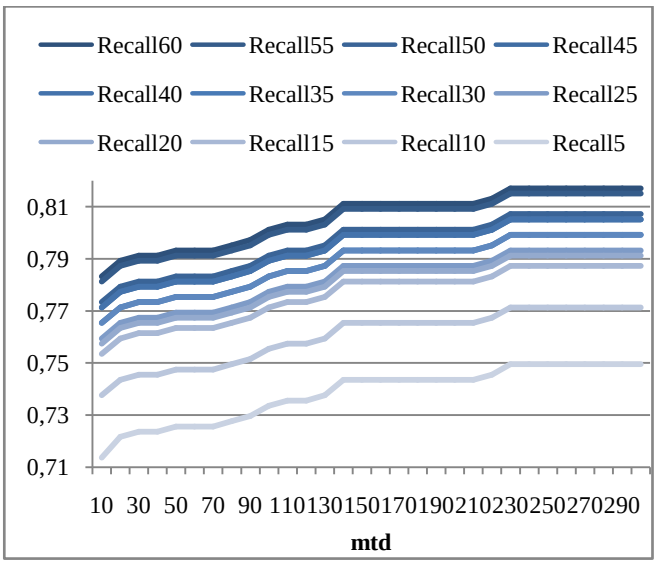



Figure 5. Recall for geotag propagation experiment in semiautomatic mode.

Considering a $t_{max}$ of 60 minutes, recall varied from 78.33% to 81.71% in semiautomatic mode, and from 72.14 to 73.97 in automatic mode.

Figure 7 and Figure 8 illustrate the comparison between precision and recall for the automatic and semiautomatic modes, for $t_{max}$ = 5 minutes (Figure 7) and $t_{max}$ = 60 minutes (Figure 8).

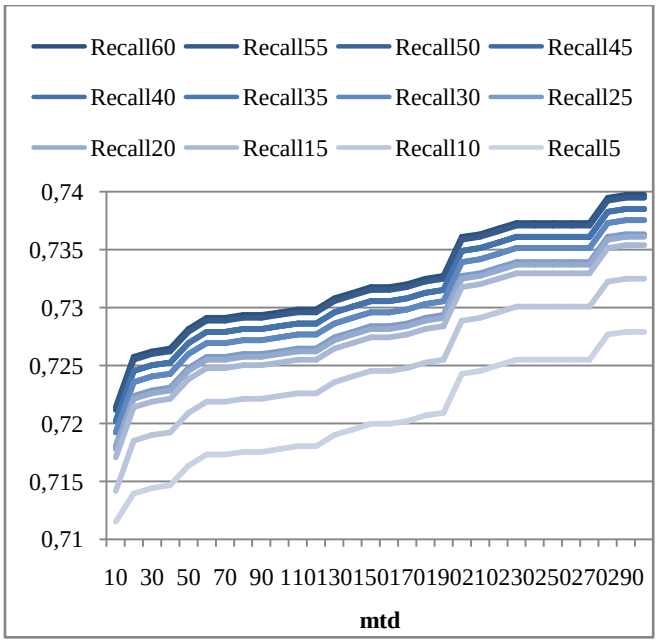



Figure 6. Recall for geotag propagation experiment in automatic mode.

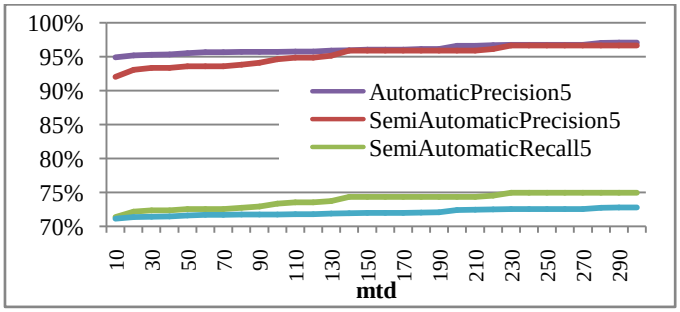



Figure 7. Comparison of recall and precision of the geotag propagation experiment automatic vs. semiautomatic, for $t_{max}$ = 5 minutes.

*B. Detection of Inconsistencies in Geotags*

To validate the geotag inconsistency detection scheme, we ran an experiment to compare the inconsistencies detected by the scheme and current inconsistencies in a database.

In order to carry the experiments out, we used a real collection with 1,040 photographs captured with a camera withan integrated GPS chip. From those photographs, 944 (90.77%) are georeferenced, among which 112 (11.86%) have the GPS status in interoperability mode. The user informed, through an application, which photographs had inconsistencies in their geotags, so that they could be compared to the inconsistencies pointed by the system, and possibly has allow the automation of the experiment.

Figure 9 presents a graphic with the precision and recall metrics. The parameter *mss* varied from 1 to 120 km/h, in steps of 1 km/h. Through the graphic, we notice that precision is directly proportional to the parameter *mss*, i.e., the higher the mean shift speed, the longer will be the maximum tolerable spatial distance between the photographs, with fewer

 ISBN: 978-1-61208-251-6 

inconsistencies detected. However, with an increasingly number of correct indications. For this reason, recall is expected to be inversely proportional to *mss*, since the number of inconsistencies pointed by the algorithm becomes smaller.

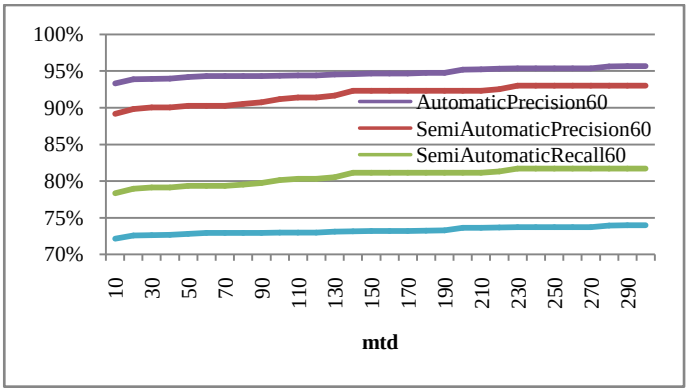



Figure 8. Comparison of recall and precision of the geotag propagation experiment automatic vs. semiautomatic, for $t_{max}$ = 60 minutes.

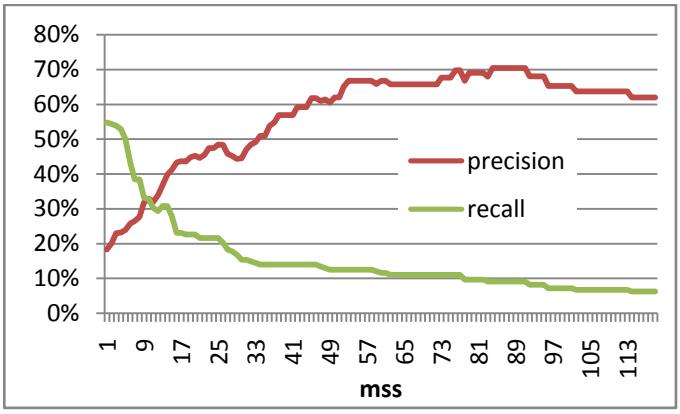



Figure 9. Graphic of precision and recall for the detection of inconsistencies.

The parameters were adjusted, and the highest precision achieved was 70.37%, for *mss* of 77 km/h. The maximum recall was of 55.81%, for *mss* = 1 km/h.

Considering a system for detecting photographs with inconsistencies, allowing the user to correct the geotags, high precision is more interesting for the algorithms, because it avoids the user from getting bored with checking too many inconsistencies erroneously reported. So, for this objective, the best value for *mss* is 77 km/h.

## II. Conclusion

In this paper, we presented a new method for detection of inconsistencies in geographic locations of photographs, and for the propagation of geotags to non georeferenced photographs.

Both approaches presented had good results. The geotag propagation achieved precision of 97.08% and recall of 73.97% in semiautomatic mode, and precision of 96.76% and recall of 81.71% in semiautomatic mode. On the other hand, the detection of inconsistencies achieved precision of up to 70.37%, proving to be a good alternative for georeference inconsistence correction schemes.

As future work, we will apply new photograph clustering methods and machine learning techniques to estimate the best input parameters for the proposed method. Besides, we will also analyze the behavior of the proposed ideas for other photograph collections.

## Acknowledgment

The authors would like to thank the CNPQ - Brazilian Research Council for funding this research.

## References

[1] H. F. Figueirêdo, Y. A. Lacerda, A. C. Paiva, M. A. Casanova, and C. Souza Baptista, "PhotoGeo: a photo digital library with spatial-temporal support and self-annotation," *Multimedia Tools and Applications*, vol. 59, no. 1, pp. 279–305, Feb. 2012.

[2] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina, "Automatic organization for digital photographs with geographic coordinates," in *Proc. of the 2004 joint ACM/IEEE conference on Digital libraries - JCDL '04*, 2004, p. 53.

[3] M. L. Cooper, "Clustering geo-tagged photo collections using dynamic programming," in *19th ACM international conference on Multimedia - MM '11*, 2011, pp. 1025–1028.

[4] K.-E. Tsay, Y.-L. Wu, M.-K. Hor, and C.-Y. Tang, "Personal Photo Organizer Based on Automated Annotation Framework," in *IIH-MSP '09*, 2009, pp. 507–510.

[5] G. M. Djuknic and R. E. Richton, "Geolocation and assisted GPS," *Computer*, vol. 34, no. 2, pp. 123–125, Mar. 2001.

[6] S. S. Lee, D. Won, and D. McLeod, "Tag-geotag correlation in social networks," in *2008 ACM workshop on Search in social media - SSM '08*, 2008, p. 59.

[7] J. Hays and A. A. Efros, "IM2GPS: estimating geographic information from a single image," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[8] L. Hollenstein and R. Purves, "Exploring place through user-generated content: Using Flickr to describe city cores," *Journal of Spatial Information Science*, vol. 1, no. 1, pp. 21–48, Jul. 2010.

[9] I. Ivanov, P. Vajda, J.-S. Lee, L. Goldmann, and T. Ebrahimi, "Geotag propagation in social networks based on user trust model," *Multimedia Tools and Applications*, vol. 56, no. 1, pp. 155–177, Jul. 2012.

[10] D. Joshi, A. Gallagher, J. Yu, and J. Luo, "Inferring photographic location using geotagged web images," *Multimedia Tools and Applications*, pp. 131–153, 2012.

[11] P.-J. Vandormael and P. Couderc, "An Algorithm to Ensure Spatial Consistency in Collaborative Photo Collections," in *SEUS '09*, 2009, pp. 332 – 342.

[12] C. Zigkolis, S. Papadopoulos, G. Filippou, Y. Kompatsiaris, and A. Vakali, "Collaborative event annotation in tagged photo collections," *Multimedia Tools and Applications*, Jun. 2012.

 ISBN: 978-1-61208-251-6 

# Integration Model for Location-Based Services in iDTV Applications

Vandecia R. Monteiro Fernandes, Geraldo Braz Junior,
Carlos de Salles Soares Neto, Anselmo Cardoso de Paiva
*Applied Computer Group*
*Federal University of Maranhão*
*São Luiz - Brazil*
*vandecia@ufma.br, geraldo@nca.ufma.br,*
*csalles@nca.ufma.br, paiva@nca.ufma.br*

Hugo Feitosa de Figueiredo,
Claudio de Souza Baptista
*Department of Systems and Computer*
*Federal University of Campina Grande*
*Campina Grande, Brazil*
*hugoff@gmail.com,*
*cdsbaptista@gmail.com*

***Abstract*—The new Digital TV (DTV) standards offer not only a significant improvement in picture and sound quality but also create the ability to run applications sent over broadcast direct to the consumers. Additionally, DTV can be watched from mobile devices, which creates a new comprehensive niche for large scale applications based on location services. This paper proposes an architectural model for integration of Location Based Services (LBS) and interactive DTV applications running on mobile devices. In particular, this work extends the middleware Ginga giving him the ability to support interactive applications based on location trough a integration layers using OpenLS to make the communication of a context server and the Ginga application. The architecture has been evaluated using real scenarios and the analysis shows promissing results.**

***Keywords*-DTV; Location Based Services; Mobile Computing; Ginga.**

## I. INTRODUCTION

Location-Based Services (LBS) are information services through a mobile network, which use geographical location to provide some specific functionality [1]. The use of user location information enables the design of innovative services, offering information increasingly accurate, precise and useful. For example, the system can provide weather information adjusted to the region where the user is or even create a free advertising-dependent environment based on locations around the user. Besides the location, other information can be monitored such as heartbeat and temperature, or environment information, such as traffic and weather condition. The information that can be collected from the situation in which the user is contained is defined as context. So context is defined as any information that can be used to characterize the status of an entity [2].

Some works have investigated the use of LBS applications and its usefulness. Among these, the LBS Framework proposed in [3] has the objective of providing mechanisms that allow the use of two separate databases as if they were a single source of information, based on the existence of two data sets: geographical and content associated to locations. However, the LBS Frawework has no location management in the mobile device and the users position is only provided upon request to the server services. FRAGIL - Framework for Information Management Localization [4] [5], is another framework that aims to provide a reusable set of location information protocols and a mechanism to allow the obtaining of location based functionalities for the development of LBS applications. Also, [6] [7] [8] present various kinds of location based applications and facilities for the development of LBS services on addressed to portable devices.

Given the mobile nature of human beings, the combination of Location-Based Services with mobile phones becomes an attractive approach for consumers and businesses. With the advent of mobile devices that allow the Digital TV (DTV) reception, the processing of the location information of devices opens doors for the development of a new niche of applications, which associate broadcast transmissions with the possibility of interactivity based on elements around the device. The interactive Digital TV (iDTV) provides to the user the ability to shop, participate in surveys, custom programming, join a class and more. The papers [9] [10] [11] present applications using a iDTV platform.

Interactivity depends on the hardware and software that are available in the mobile device. The coordination between hardware and software is done by a software layer called middleware. The middleware adopted as the standard for the Brazilian DTV, in accordance with specifications of the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) for IPTV is called Ginga [12]. The main function of the Ginga middleware is to ensure interoperability in the development of iDTV applications, providing a standardized virtual machine on which to run interactive applications.

This paper proposes an integration model for Ginga applications and Location Based Services (LBS) on mobile devices, extending the range of applications and information that could be treated by users on the move. Specifically, the proposed architecture aims to provide mechanisms of self-triggering using the geographic context for mobile applications on the middleware Ginga. As a result, iDTV applications may be extended to location-aware capabilities.

 ISBN: 978-1-61208-251-6 

The remainder of this paper is organized as follows. Section II focues on the background concepts needed for the proposed architecture presented in Section III. In Section IV, we address usage scenarios of the architecture from the perspective of mobile digital TV. Finally Section V concludes the paper and discuss further work to be undertaken.

## II. BACKGROUND

This section presents the theoretical background needed for the construction of the proposed architecture. Subsection A presents the VadeMecum framework, used to manage the information of user's context. Subsection B focuses on details about the Ginga middleware, which is extended by this work to support Location Based Services. And, the final subsection presents the OpenLS, that aims at promoting interoperability for the proposed architecture.

### A. VadeMecum

It is common in context based applications to have one module or component responsible for managing rules and actions that must be taken whenever a particular client has satisfied given conditions.

The VadeMecum uses a model-based inference rules to monitor the contextual states of users in the system. Thus, you can add rules that will launch actions if these are met. A rule in the VadeMecum is like E-C-A (Event-Condition-Action), in which the event is to update some contextual information monitored; the condition and description of a state contextual and action is the operation to be performed when the condition is satisfied. For example, a possible rule is: when you have some close contact geographically, this location should be shown on the map on the mobile device user. A user can have multiple registered rules and can either activate or disable them at any time.

In this proposal, the component that performs this task is called *ContextServer*. Its basic functionality is to manage a database of rules and actions previously registered. As the client application is used, a context database is fed and accessed by the *ContextServer* that checks whether any rule or action should be taken according to the context in which the customer is located. A server implementation of the *ContextServer* is VadeMecum, formed by the tool CARE (Context Aware Rule Editor) and the server itself, proposed in [13] and visualized in Figure 1.

This server is responsible for storing context, perform inference and monitor contextual information. Its operation is based on contextual rules implemented using ontologies, which indicate what actions should be implemented in applications when a particular contextual state is reached. The addition of the rules in the server is performed by the end user through the CARE tool that assists the user in the process of specifying such rules and provide a lightweight

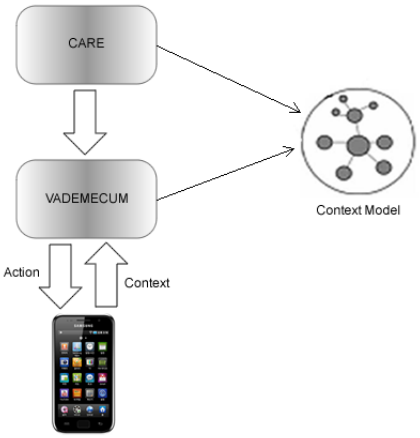



Figure 1. Context Server VadeMecum

and intuitive graphical user interface, so that it is easy for users to express their rules.

After creating the rules, using the VadeMecum, they should be sent to the context server that will monitor them along with the state of the contextual user. The storage and retrieval of contextual information is made via a context model that logically describes how information is stored, following the specification of an ontology created previously. The insertion and retrieval operations are performed by a communication protocol that follows the HTTP (POST and GET) and languages SPARQL and SPARQL Update [14]. In order to monitor the server state there is a contextual inference engine based on rules. Once the rule is activated, this will be accessed every time that the subset contained in the contextual state condition is changed. When the condition is satisfied, the action is executed and the rule will be disabled until the contextual state is no longer valid in the condition.

In the VadeMecum there are three possible actions:

- ShowMultimedia that receives as a parameter the user who receives the action and a multimedia file,
- ShowOnMap that receives as a parameter the user who receives the action and an object that will be shown on the map location,
- SendMail, that receives an e-mail and sends a message to the recipient.

However, it is possible that the actions in VadeMecum trigger external services so they send some action itself, which may be executed in a user's mobile device.

### B. Ginga Midleware

Ginga middleware is the name of the Japanese-Brazilian Terrestrial Digital TV System and ITU-T standard for IPTV. The Ginga middleware was developed entirely in Brazil providing two types of programming environments: a

 ISBN: 978-1-61208-251-6 

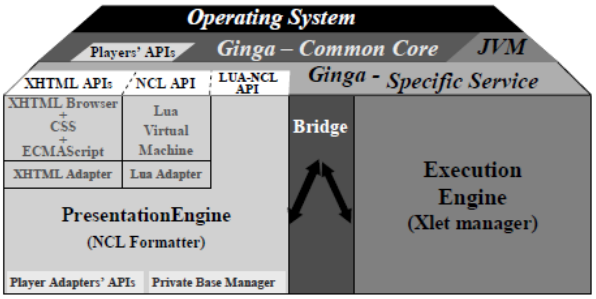



Figure 2. Ginga Archicteture [15]

declarative, represented by Ginga-NCL, and an imperative, represented by Ginga-J [15].

Ginga-NCL uses a declarative language based on Nested Context Language (NCL). Through this language, an author can describe the temporal behavior of a multimedia presentation, associate hyperlinks (user interaction) to media objects, define alternatives for presentation (adaptation) and describe the layout of the presentation on multiple devices. The Ginga-J, is a logical subsystem of the Ginga system responsible for processing active content. A key component of the execution environment is Ginga machine for implementing the mandatory content, comprising a Java virtual machine.

The architecture (Figure 2) and Ginga facilities are designed to be applied to broadcasting systems and terrestrial broadcasting receivers. Additionally, the same architecture and facilities can be applied to systems using other data transport mechanisms (such as satellite systems and cable systems or IPTV). For more information about Ginga, its applications and models see [15].

### C. OpenLS

Interoperability is one of the key points to be considered in developing location-based systems, since these should be available on different platforms and operating systems. The OpenGIS Consortium [16] defines a set of computational patterns that aim to promote interoperability between Geographic Information Systems (GIS).

The GIS automated systems are used to store, manipulate and analyze geographic data. In the 80s, with the advent of personal computing, there has been a rapid development of these systems, as well as for military use and environmental planning. Since then, GIS has evolved to make technologies most effective and affordable.

The OpenLS specification [17] was adopted by the OpenGIS Consortium, and is focused on the development of interface specifications that facilitate the use of location and other forms of spatial information about the environment of wireless Internet. The goal of this initiative is to specify a set of interfaces, standards and protocols on which developers can use to integrate geospatial data and geoprocessing resources into location services and telecommunications infrastructure, providing these capabilities for a variety of applications.

The OpenLS is used to allow access to mechanisms for location of the network and a set of services called OpenLS Core Services. The specified services are Location Utilities Service, Directory Service, Presentation Service and Route Determination Service. The applications use the interfaces of OpenLS to get the content needed to perform their functions. Such content would be: data maps, route networks, addresses, navigation information, directories of places with information, products or services.

## III. THE GINGA-LBIA ARCHITECTURE

The proposed architecture, named Ginga-LBiA, aim to integrate Location Based Services and iDTV application under the Ginga middleware. The architecture is divided into three modules: Context Server, Gateway and Client, as shown in Figure 3 below.

The Context Server manage rules and actions to be analyzed according to the DTV client context. This module includes the Client Service Listener, Service Context Notify application and mobile digital TV. To link the two platforms: server and application Context Digital TV, we implemented a translator module represented using the Gateway. Each module and its composition are described in more detail in the following subsections.

### A. Service Context

This component aims to store and perform inference rules based on the user contextual information and to propose an action to be taken through the VadeMecum. However, VadeMecum was extended to communicate with the Ginga Digital TV. An important aspect of the proposed architecture is that it can be easily extended by other server implementations of context. To obtain a model that is implementation independent, it is necessary to create an interoperability layer. The same is proposed in this architecture by implementing a communication layer, independent of context server, using OpenLs as the standard protocol for information exchange. This component comprises the Gateway.

### B. Client Module

The client module consists of three components: the DTV application, Service Listener and Service Context Notify, as presented in Figure 3. The application runs on a mobile device, emulating the Ginga middleware, according to the proposed implementation defined in [18]. It is necessary to extend the same support to the management of context information such as the geographical location. Thus, the application itself, through the Ginga middleware, informs the device location. The Notify Context Service Component is the component responsible for periodically updating data

 ISBN: 978-1-61208-251-6 

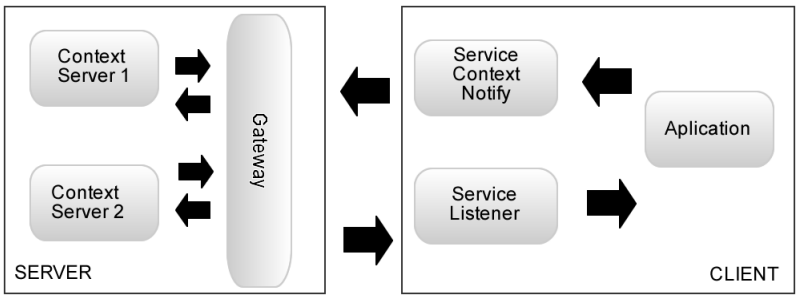



Figure 3. Proposed Architecture: the Ginga Location Based iDTV Architecture (Ginga-LBiA)

in the context server with client information. To accomplish this task, the Service Context Notify receives the information from the application and sends it to the Gateway module according to some update policy to be defined.

Once the context server evaluates the information given by the previously created rules, if there is any action to be taken, it forwards the action to the application through the Gateway explained on next section. The component responsible for managing the actions received by the server context is the Service Listener. If the action is accepted, it is forwarded to the application where an event is triggered for the user.

### C. Gateway

The Gateway module is responsible for communication between the server and the client module context. It is implemented according to the OpenLS specification [17], which defines service interfaces that facilitate the development of location-based applications. In the OpenLS specification, the service request occurs via a SLIR (Standard Location Immediate Request). The result obtained is sent to the user via a SLIA (Standard Location Immediate Answer). The Gateway has different functions according to the type of architecture that is used. In the model which uses WebServices, the request has to be dealt with SOAP technology. If the model is adopted using the server VadeMecum, which uses a model-based inference rules, the Gateway will have to turn the request of the application in a triple, such as subject, predicate, action, using a language as SPARQL [14].

## IV. Scope of the Proposed Architecture from the Viewpoint of Digital TV Applications

Figure 4 presents a common scheme for application distribution on interactive digital TV. The TV broadcaster transmits to every device the same application multiplexed with the main video and audio. This interactive application can be received by fixed terminals in either set-top boxes or mobile receivers. In both cases, the applications are running on top of a middleware layer. In this content distribution environment , applications can also provide information back to the TV station using the Internet as a communication

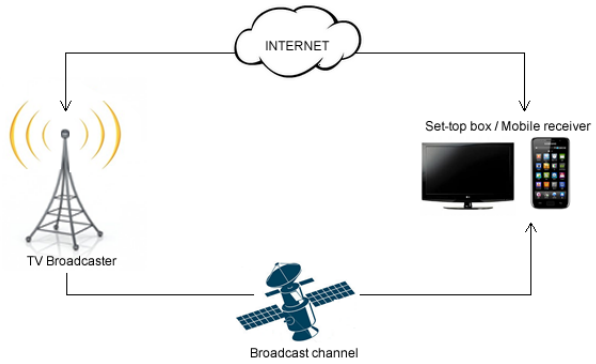



Figure 4. Common scheme for applications distribution over interactive digital TV

channel, which allows that information about the users are collected.

The components of the Ginga-LBiA to support location service can also be distributed among environments and TV broadcaster mobile device. There are three main ways on how this architecture can be distributed, which characterizes the supported location application profile being performed.

*1) Fully Broadcast Applications:* In this category of applications, all message exchanges occur in the client that acts only like a receiver. It is not necessary to have return communication channel to the TV broadcaster, which makes this category the most common. The iDTV application is transmitted to all viewers by the broadcast channel. It is through this same channel that the entire contents of the application is received on the client side. This means that all alternatives that may be displayed depending on the location must be communicated to all customers alike. Figure 5 shows the architecture that uses only the broadcast transmission.

As an example of possible fully location-based broadcast applications is a scenario of location based advertising. In this category, there is no request of content made by the application, it only receives content. Suppose a user walking in a shopping center. In this case, the station can transmit an specific advertisement just to the users that are near that mall. Thus, the application is adapted according to

 ISBN: 978-1-61208-251-6 

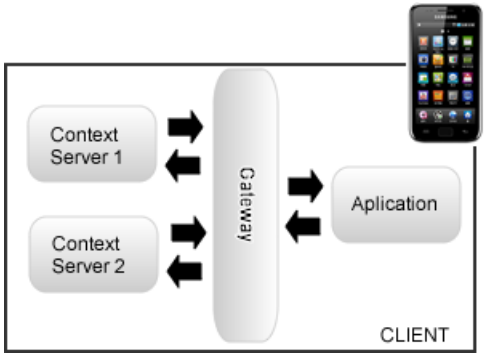



Figure 5. Fully broadcast applications using Ginga-LBiA architecture

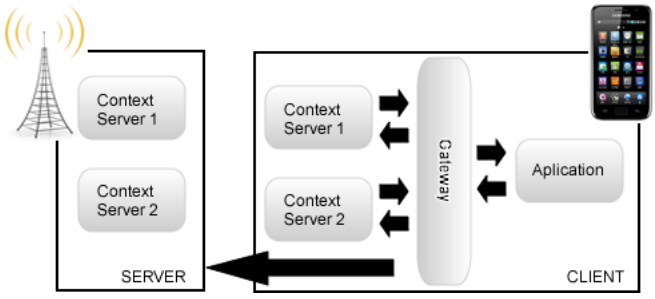



Figure 6. Illustration of message exchanges for hybrid application using Ginga-LBiA architecture

the location. In other words, the advertisement is presented depending on whether the user is close or not to the mall. The same scenario can be applied, for example, on public transportation such as subways and buses. If the tube is equipped with a TV running the application of digital TV, when it pass through certain areas, the application will conveying advertisements about those areas.

*2) Hybrid Applications:* In hybrid applications, the content traffic has the jointly responsibility in client and server. The iDTV application will be running on the client, and when asked for additional content, will communicate with the server to answer the request. Figure 6 presents a illustration of message exchanges of the architecture in this category.

*3) Applications without Broadcast:* In this category only iDTV application is resident on the client, all requests are answered by queries to the server context. In this category we can cite tour guides. In tourist towns, for example, tourists can see detailed information about buildings and historic sites. In the user mobile device, just the iDTV application was running. When the user moves, the application will pose a query to a server and receive audiovisual content specific to the location where he is. Figure 7 exemplify this kind of application.

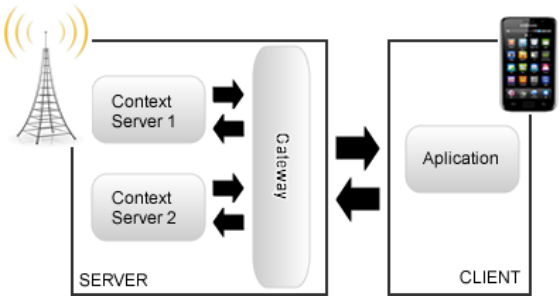



Figure 7. Illustration of a application without broadcast using Ginga-LBiA architecture

## V. Conclusion

Digital TV is an important mechanism for transmitting information. Television is known as the principal communication vehicle that reaches all the social classes and regions of the world. Interactive applications in Digital TV allows for the implementation of further tools, regardless of location or culture.

By joining DTV with Location Based Services it is possible to enhance the interaction through regionalization solutions, which provides a custom activity and content. Thus, applications on advertising, social media or government could be extended to deal with specific and localized cases.

With this goal, this paper presented the Ginga-LBiA architecture for integration of location based applications on the Ginga middleware. The integration was performed using a methodology that uses the OpenLS standard in order to achieve interoperability between different applications and servers.

The architecture was analyzed on the main scenarios of applications in DTV, which correspond to the following categories: applications with or without broadcast and hybrid. In all categories, the architecture was adequate to provide appropriate services independently of application kind. The proposed architecture treats location features in a transparent manner and provides interoperable and adaptable communication mechanisms. Thus, this architecture could also be applied for others midlewares DTV just changing Ginga midleware by the appropriate one. This could be reached because all the communication interfaces are based on OpenLS and the Gateway layer is independent of the midleware choosen.

Finally, this approach shows promising to integrate Digital TV services with location-based services. With this architecture, it becomes possible to build interactive applications that can capture the context of the client to provide mechanisms to adapt the content appropriate. However, it is understood that there are other context information beyond the location that could be handled by the architecture, such as wheather conditions, temporal and

 ISBN: 978-1-61208-251-6

social contexts. As further work, we intend to incorporate new extensions to the Ginga to incorporate this contextual information. We also intend to develop security and privacy features as well compare this architecture with other that would exists to capture the keys advantages and disavantages and provide a better interoperability to our approach.

## ACKNOWLEDGMENT

The authors would acknowledge CNPq, CAPES and FAPEMA for the financial support.

## REFERENCES

[1] R. Ferraro and M. Aktihanoglu, *Location-Aware Applications*. Manning Publications Co., 2011.

[2] P. Prekop and M. Burnett, "Activities, context and ubiquitous computing," *Computer Communications*, vol. 26, no. 11, pp. 1168–1176, 2003.

[3] K. Andersen, M. Cheng, and R. Klitgaard, "Framework for building location based services," 2003, technical report, Aalborg Universitet.

[4] A. de Paiva, E. Monteiro, J. Rocha, C. de Souza Baptista, A. Silva, and S. da Rocha, "Location information management in lbs applications," *Encyclopedia of Information Science and Technology*, pp. 2450–2455, 2009.

[5] E. Monteiro, A. de Paiva, and C. de Souza Baptista, "Arquitetura de um framework para o desenvolvimento de aplicações baseadas em localização," in *Conferência IADIS Ibero-Americana WWW/Internet*. IADIS, 2005, pp. 336–343, in Portuguese.

[6] N. Fernando, D. Dias, and S. Wijesekara, "A framework to develop location based services applications using ogc map services," in *Information and Automation for Sustainability (ICIAFs), 2010 5th International Conference on*. IEEE, 2010, pp. 521–526.

[7] A. Desai, Z. Laliwala, and S. Chaudhary, "Context and location based service grouping and group notification," in *Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific*. IEEE, 2009, pp. 227–232.

[8] S. Kumar, M. Qadeer, and A. Gupta, "Location based services using android," in *IMSAA09: Proceedings of the 3rd IEEE International Conference on Internet Multimedia Services Architecture and Applications*, 2009, pp. 335–339.

[9] M. Ferreira, M. Braga, P. Santos, and N. Santos, "An application model for digital television in e-learning," in *Interactive Computer Aided Blended Learning-ICBL Conference. Florianópolis:[sn]*, 2009.

[10] V. Silva, R. Rosa, and V. De Lucena Jr, "Using brazilian digital tv to integrate health care services embedded in medical commercial devices," *Medical Informatics*, vol. 1, no. 1, pp. 93–110, 2012.

[11] O. Maia, N. Viana, and V. de Lucena Jr, "Using the idtv as the center of an ubiquitous environment," *Medical Informatics*, vol. 1, no. 1, pp. 226–248, 2011.

[12] L. Soares and G. Souza Filho, "Interactive television in brazil: System software and the digital divide," in *European Conference on Interactive TV (EuroITV), Áustria*, 2007.

[13] H. F. Figueiredo, "Uma infraestrutura de suporte a aplicações cientes de contexto com o enfoque no usuário final," Master's thesis, Universidade Federal de Campina Grande, Campina Grande, 2008, in Portuguese.

[14] W. Consortium, "W3c recommendation sparql," 2008, available at http://www.w3.org/TR/rdf-sparql-query, last accessed Ocotber 21, 2012.

[15] L. Soares, R. Rodrigues, and M. Moreno, "Ginga-ncl: the declarative environment of the brazilian digital tv system," *Journal of the Brazilian Computer Society*, vol. 12, no. 4, pp. 37–46, 2007.

[16] O. Consortium, "Opengis standards," 2012, available at http://www.opengeospatial.org/standards, last accessed Ocotber 21, 2012.

[17] I. Open Geospatial Consortium, "Opengis location services (openls): Core services [part 1-directory service, part 2-gateway service, part 3-location utility service(geocoder/reverse geocoder), part 4-presentation service, part 5-route service]," 2008, available at http://www.opengeospatial.org/standards/ols, last accessed October 21, 2012.

[18] G. Ferreira, G. Nogueira, G. Comarela, F. Fabris, M. Martinello, and J. P Filho, "Ginga-ncl em dispositivos portáteis: Uma implementação para a plataforma android," 2010, in Portuguese.

 ISBN: 978-1-61208-251-6 

# Starting a Spatially Enabled Society: a Web System for Collecting Volunteered Geographic Information in the Area of Public Security

Jarbas Nunes Vidal Filho, Jugurta Lisboa-Filho, Wagner Dias de Souza, Diogo Franklin de Oliveira
Universidade Federal de Viçosa (UFV)
Departamento de Informática (DPI)
Viçosa – Minas Gerais - Brazil
{jarbas.filho,jugurta}@ufv.br, {wagner.supremo, diogofranklinn}@gmail.com

***Abstract*— A spatially enabled society provides to citizens, businesses and government access to spatial data regardless of computing platform. The article describes a collaborative Web system to gather geographic information about the area of public security and disseminate it at all levels of a society using Web 2.0 tools. The data generated by the user become another source of data for agencies involved in public security to use the information in order to try to minimize violence and allow citizens to protect themselves from criminality.**

***Keywords- Volunteered Geographic Information; Web Systems; Google Maps; Public Security.***

## I. INTRODUCTION

Currently, the use of geographic information is fundamental to sustainable development as well as improvements in the making decision process by a particular organization, in which the factor "where" is important and precedent to perform an action on several sectors, such as public security, health, urban planning, deforestation control, agricultural production and others. The economic development, social stability, improvement of citizen's public security, and others can be reached by developing products and services that are based on geographic information collected by all levels of a society or government [1].

These actions can be facilitated by the development of a spatially enabled society [1]. In this type of society, spatial information is regarded as common goods made available to citizens and organizations to assist in developing techniques to support decision making [1, 2]. A spatially enabled society in the context of public security may be useful to bring citizens closer to police departments and improve the activities within these organizations. Any user may search for a service or information, which should always be available and reliable access [3]. Moreover, people can voluntarily contribute information to help security departments to make decisions based on volunteered information.

For the development of such society, it is necessary the use spatial data. These, in turn, need to have a certain degree of quality, sufficiently reliable, easily accessible and available in real time for use by that society [1]. Based on this understanding, Spatial Data Infrastructures (SDI) are being developed in many countries as a platform that improves the sharing, access and integration of spatial data and services. An SDI can also store spatial data and provide services of access to data supplied voluntarily.

In recent years, the amount of spatial data generated by the users has grown. Increasingly users are indexing content based on their geographic coordinates. With this new paradigm of spatial data production, the term Volunteered Geographic Information (VGI) has arisen, identified by Goodchild [4] as a new phenomenon to describe user-generated information, combining elements of Neogeography, Collective Intelligence and Web 2.0 [5].

It is possible to note an evolution in the treatment and production of spatial information. With the popularization of the Internet and the emergence of Web tools, it is possible to observe a breakthrough in the development of new technologies for the production of geographical information [4]. Considering that Web 1.0 initially assumed a unidirectional role only allowing users to visualize information, Web 2.0 plays a bidirectional role in which users are able to interact and provide information that can be accessed by any user who has access to the computer network.

The purpose of making a society become a spatially enabled one is to transform the citizen into a "voluntary human sensor" in the Web 2.0 world, providing information and reviewing information that has been contributed by other voluntary users. According to Georgiadou et al [6], the use of "participatory human sensors" can extend the power of action of ordinary citizens, improving the ability of citizens to directly influence the services performed by a particular organization.

The rest of the paper is organized as follows. Section II describes the main related work. Section III presents the motivations and objectives of the project and describes the case study developed. Section IV describes the methodology used and some conclusions are presented in Section V.

## II. RELATED WORK

There are many works developed with the use of information technology and employing spatial data, applied in the area of public security. However, these works are ceasing

 ISBN: 978-1-61208-251-6 

to be developed in a conventional environment of Geographic Information System (GIS) and starting to use Web 2.0 concepts and neogeography to produce data from ordinary citizens, i.e. without knowledge of specific areas of geosciences. Some examples of these projects are listed below.

The project Wikicrimes [7] main idea is the access to information and collaborative mapping of crime in a digital map using the Google Maps API. Any user with a mobile device or a computer with Internet access can cooperate with the system and seek information about a registered crime. This work contains functions to automatically validate registered crimes, display statistics of them and identify hot points of the city, i.e., identify areas of greatest risk, using an algorithm based on Kernel Map.

The project called "Paz tem voz – Mapa do Crime" [8] or "Peace has voice - Crime Map" in English is a project developed to collect volunteered information from police reports that the citizen was a victim or witnessed the fact. The system consists of a module of voluntary collaboration, functions for visualization of statistical analysis of recorded data and identification of homicide profile, such as age, crime motivation and weapon used. Moreover, it is also possible to make a complaint about theft, burglary and other crimes. The system uses the Google Maps API and is designed to meet the desires of the people of the state of Paraná, in southern Brazil.

The project CrimeViz [9], developed at the University of Pennsylvania, also uses the Google Maps API to provide digital maps of Washington, District of Columbia, United States. However, this study differs from others cited above in two aspects. Firstly, the CrimeViz is fed by official data of the District of Columbia and not by voluntary contributions. Secondly, it implements a panel of spatio-temporal analysis by day, month or year. This panel can be understood as a box containing a button to play, pause and continue. The user selects the unit of time (day, month or year) and selects the play button. Then, the system will display a color histogram representing the number of incidents that were reported and it is possible to view on the map a spatio-temporal analysis showing the data records in the last days, months or years.

The works cited above have great contributions to public security in combating violence. The work proposed here differs by proposing some contributions that are useful for approximating the population to police agencies by, for example, using concepts of Wiki in VGI data in order to make the citizen a reviewer of user-generated content. Another example is the development of a forum to discuss the recorded data, creating a collective intelligence that can be useful to perform validation of data, from the testimony of others or serving as collection for more information about a registered crime. Furthermore, in order to improve access to and dissemination of VGI data, an SDI is used for data storage. Finally, the potential of VGI in the area of public security can be validated based on some statistical analysis on the VGI data in relation to official bases of police agencies.

## III. MOSSORÓ SPATIALLY ENABLED: MOTIVATIONS AND OBJECTIVES

The city of Mossoro, located in western Rio Grande do Norte state, in Brazilian northeast, is the second most populous city in the state and has a large growth related to industries and companies that have settled in the city such as Petrobras. In recent years the feeling of insecurity in the city has grown among citizens. The "Instituto Técnico-Científico de Polícia (ITEP)" or Scientific-Technical Institute of Police, an institute responsible for statistics from police reports, started to show alarming numbers in recent years related to the increase in crime in the city.

According to data from the Sangari Institute [10], the number of homicides in Mossoró city has been growing year by year. In 2009, 132 homicides were recorded. The number rose to 178 homicides in 2010 and to 196 homicides in 2011. In 2012, until September, 97 occurrences of homicide had already been recorded. Due to this alarming increase, the case study of this work is related to Mossoró city.

The aim of this study is to make the citizen a participatory "voluntary human sensor" in a spatially enabled society using Web 2.0 tools. Any user who uses a mobile device or a computer with Internet access can report problems of insecurity or register crimes in which the citizen was the victim or a witness. The data generated by the user are stored in a database integrated with a SDI, making it easier to search for data by the various organizations that have some direct or indirect influence in the area of public security and the dissemination of data at all levels of society, as well as avoiding efforts by users in creating repeated data. The user can record VGI data, participate in the forum of a registered data in order to enhance the data or provide other relevant information, and use Wiki services to review data from voluntary contributions, as it is done at the free encyclopedia, Wikipedia.

## IV. MATERIALS AND METHODS

The collaborative systems can be understood as tools that support Web 2.0 resources and allow the direct or indirect interaction of a particular individual or of collective groups with any content available on the Internet [11]. This type of system emerged in recent years due to the advances in Web. Until recently, the Web was used only to provide information, making users only consumers of information. This phase is known as Web 1.0. With advances in technology it has become possible to interact with content published on the Internet. Blogs, Wikipedia, Email and others are examples that allow users to interact with content, setting up another phase of the Web known as Web 2.0. The system developed in this work uses technologies that support Web 2.0 capabilities and other tools needed for development, such as: (1) Apache; (2) Google Maps API version 3 for map viewing, statistical analysis and analysis of the most dangerous areas based on the algorithm of Kernel Map; (3) The programming languages Hyper-Text Markup Language (HTML), Cascading Style Sheets (CSS), Hypertext Preprocessor (PHP), AJAX, jQuery, JavaScript and eXtensible Markup Language (XML) were used for customization of the Google Maps API, development of collaboration module, interface and other features; (4) The Database Manager System MySQLServer is used to store

 ISBN: 978-1-61208-251-6 

users' contributions and PHPMyADMIN is used as database configuration assistant.

The Apache server (or HTTP Apache server) is a free software developed by the Apache Software Foundation, and aims at processing information on the Web, i.e., it is a server responsible for answering Hyper-Text Transfer Protocol (HTTP) requests, Web standard protocol [12]. The system is available in the domain (www.ide.ufv.br/mossorocrimes) using a machine with Internet access as the physical server and the Apache as Web server.

This work used the Google Maps API v3 with support for browsers (Internet Explorer, Firefox, Safari, Opera and Chrome) and mobile devices with Android and Iphone. This API provides several features for viewing 2D and 3D maps, geocoding services, best route and others. Features of this API are used to perform statistical analysis of VGI data and risk analysis, i.e., determining the most dangerous areas in the city. To use the API it is necessary to possess a valid license key and provided by Google itself. This key is used together with JavaScript codes that can be loaded into the system. Google provides all documentation for use of the API and the resources supplied by it.

The collaboration module is a service offered by the system to collect VGI data. The user can provide textual information about an act of violence, insert videos and photos related to the registered act and use the forum to discuss the data, providing relevant information to help improve the quality of data. This module was developed with the programming languages PHP, JavaScript, HTML and XML, supplying the user with greater interaction to insert, update or criticize a VGI data.

Data collection is an important part of this work, because from them, it is possible to perform statistical analysis of VGI data and possibly infer about the potential of the VGI. Therefore, it is necessary to create a database allowing the authors to hold VGI data instead of them going to the Google server. The MySQLServer was used due to it being easy to install and use, being free software and having the PHPMyADMIN assistant that facilitates the creation, modification and settings of the database.

## V. RESULTS AND DISCUSSIONS

The developed system consists of a prototype to collect VGI data and perform statistical analysis of VGI data and by geographic region. The software was developed to support requests coming from computers (laptop, desktop) and smartphones that support Android or Iphone operating systems. Fig. 1 shows the system interface that uses only HTML and CSS in its development to make viewing very simple in devices that have limited screen such as mobile devices.

The system provides two categories: security and others. At the time the user performs collaboration he can tell which category and type of occurrence that he will provide. The category "other" represents other data that have some connection with the area of public security, for example, report lack of lighting on a street. Fig. 2 shows statistical analysis on the collaborated data by categories and on users who cooperated more. Then it is possible to have a control over the most active users and filter, in the database, the information of these users in order to try to identify whether the user is collaborating to help or disturb police activities.


Figure 1. Interface of the system.

The identification of risk areas is essential when the topic is public security. The Google Maps API provides resources to implement services for risk analysis based on Kernel Map, which is a statistical method of estimation of curves, i.e., is plotted on the map using interpolation methods, the intensity of a particular phenomenon in the region [13]. The system has a button called "heatmap" to enables the function that implements the Kernel Map. Moreover, as the user moves the map it is possible to view the statistics of VGI data by region. Fig. 3 shows the use of Kernel Map in recorded incidents.

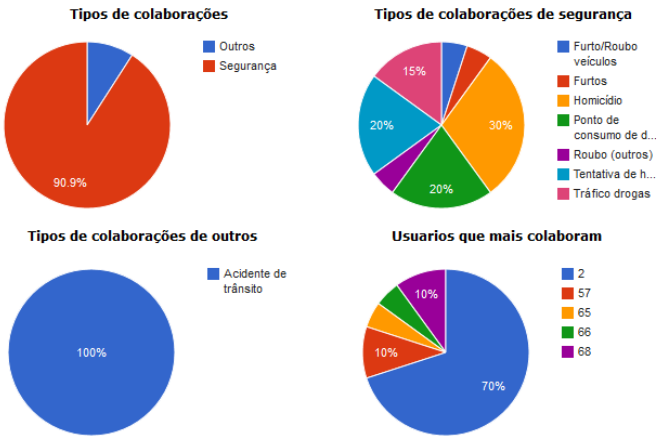



Figure 2. Statistical analysis by category and user.

Because there are few collaboration, it is not possible to infer any relevant result on the Kernel map. From the moment that there is more collaboration in the system it will be possible to see the color intensity by region.

The collaboration module is responsible for receiving contributions from users. To perform a collaboration is necessary that the user is "logged in" to the system so it is necessary a simple registration. This information is preserved and is not available at any time that the user is interacting with the system. After accessing the system using the credentials of the registration, the user can click in the “collaborate” tab and then click on the map to write a collaboration. The user can register and send textual information, photo, video or any other file. Fig. 4 shows the "infowindow" of collaboration module with the tabs to send multimedia files. Mossoró has Streetview service. The citizen can use this service to better identify the place for registration of the occurrence.

 ISBN: 978-1-61208-251-6 


Figure 3. Risk analysis based on Kernel Map.

Once the user registers the collaboration by clicking on "send collaboration" it is created a forum about the data. To interact with the forum, the user must be "logged in" to the system. Therefore, any registered user can interact with other users and provide extra information that was not addressed in the description of the occurrence. This generates a collective intelligence about the data, similar to those that exist in relevant comments in posts on social networks, whose goal is to help in the description of a photo, text, video or other.

The system also provides access to all data recorded and it is possible to follow the updates in real time and analyze all data that have contributed to the system. Moreover, it has a data filtering service whose goal is to facilitate the viewing of specific data such as homicides, thefts and others. These data can be filtered and analyzed separately using Kernel analysis.

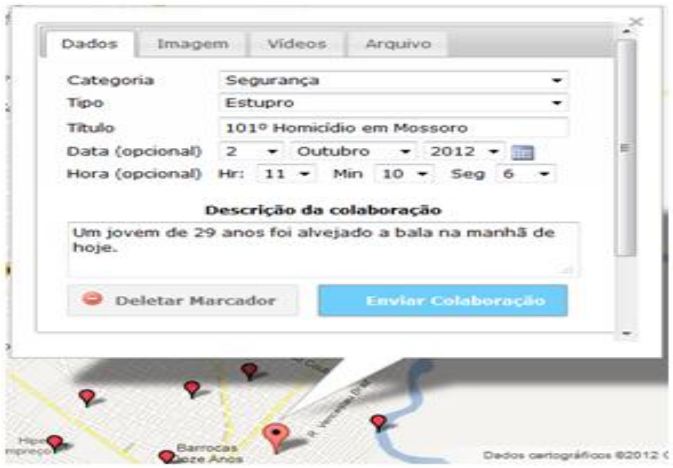



Figure 4. Infowindow of collaboration module.

## VI. Conclusion and Future Works

A collaborative Web system is essential for citizens to share information that are often available on websites that do not have the knowledge of agencies in charge of public security. A society enabled with VGI data on public security can be useful for companies, citizens and government using another database to improve the decision-making process. VGI data being used at a local level, which is highly recommended, can provide a faster response to security departments than their own official data. Therefore, in an emergency situation it is possible to get better results in the execution of activities.

This type of system when ethically used can provide a satisfactory database for agencies responsible for security in a city and help people to protect themselves from violent places. In future work a Wiki module will be developed so that citizens can interact in order to improve VGI data quality and an SDI that will provide access to all the agencies. So at major events in the city, the agencies that are related to public security can use a database containing information from other sectors and improve their activities in combating violence. To validate the VGI data, statistical analysis will be conducted comparing the VGI data with official data to assess the impact of VGI in security.

## Acknowledgements

This project was partially funded by CNPq, Fapemig, CAPES, Funarbe and the Company Sydle.

## References

[1] A. Rajabifard, “A Spatial Data Infrastructure for a Spatially Enabled Government and Society”. In A multi-view Framework to Assess Spatial Data Infrastructure. Austrália: Digital Print Center, 2008, pp. 11-22.

[2] H. Mohammdi, A. Colebatch, G. Dawson, and S. Ballingall, “Intelligent Speed Assist: Spatially enabling societies”. In Proc. of GSDI, Singapore, 2010. Available from: <http://www.gsdi.org/gsdiconf/gsdi12/papers/73.pdf>.

[3] I. Williamson, A. Rajabifard and J. Wallace, “Spatially Enabling Government: An Internacional Challenge”. Available from: <http://gsditest.opengeospatial.org/gsdiconf/gsdi10/papers/TS2.3paper.pdf>. Access: 01 September 2012.

[4] M.F. Goodchild, “Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0”, *International Journal of Spatial Data Infrastructures Research*, v. 2, pp. 24-32, January 2007.

[5] A. Hudson-Smith, A. Crooks, M. Gibin, R. Milton and M. Batty, “NeoGeography and Web 2.0: concepts, tools and applications”, *Journal of Location Based Services*, v. 3, pp. 118-145, June 2009.

[6] Y. Georgiadou, B. Bana, R. Becht, R. Hoppe, J. Ikingura, M. J. Kraak, K. Lance, R. Lemmens, J.M. Lungo, M. Mccall, G. Miscione, and J. Verplanke, “Sensors, empowerment, and accountability: a Digital Earth view from East Africa”, *International Journal of Geographical Information Science*, v.4, pp. 285-304, July 2011.

[7] Wikicrimes. Available from: <www.wikicrimes.org>. Access: 20 September 2012.

[8] Paz tem voz – Mapa do Crime. Available from: <http://www2.gazetadopovo.com.br/pazsemvozemedo/mapadocrime/>. Access: 22 September 2012. (In Portuguese)

[9] CrimeViz. Available from: <http://www.geovista.psu.edu/CrimeViz/DC/v2/>. Access: 22 September 2012.

[10] Instituto Sangari. Available from: <http://www.institutosangari.org.br/instituto/>. Access: 25 September 2012. (In Portuguese)

[11] V. Furtado, L. Ayres; M. Oliveira, E. Vasconcelos, C. Caminha; J. D’Orleans and M. Belchior, “Collective intelligence in law enforcement the wikiCrimes system”, *Information Sciences*, v.180, n.1, pp. 4-17, January 2010.

[12] The Apache Software Foundation. Available from: <http://www.apache.org/>. Access: 25 September 2012.

[13] An Introduction to Kernel Methods. Available from: <http://intellisysdev.enm.bris.ac.uk/cig/pubs/2000/svmintro.pdf>. Access: 04 December 2012.

 ISBN: 978-1-61208-251-6 

# Automated Processing of Postal Addresses

Konstantin Clemens
*Deutsche Telekom Laboratories, TU Berlin*
*Service-centric Networking*
*konstantin.clemens@campus.tu-berlin.com*

***Abstract*—Postal addresses are involved whenever post mail is to be delivered to an addressee, or to describe the location of a person or organization. The world-wide adoption of postal addresses and the lack of a centralized control entity led to very heterogeneous postal address formats. Even within a single country various postal address formats may be used. In this paper the main tools for processing postal addresses in an automated fashion are compared: Multidimensional indices, address patterns and gazetteers. Special focus is put upon supporting as many heterogeneous postal address formats as feasible with the various tools.**

***Keywords*-postal automated processing; postal addresses; address patterns; gazetteer; spatial indices**

## I. INTRODUCTION

Postal Addresses identify addressees so that no room for ambiguity is left. There was no need for that before modern postal systems began offering their service to the broad public. Until every citizen was eligible to receive post office mail, many regions lacked house numbers or street names. That still is the case in many areas. E.g., in Japan often named houses are used over numbered houses on named streets. Extremely poor neighborhoods without official organization as the Brazilian favelas often lack street names and house numbers too.

To disambiguate the name of an addressee, postal addresses utilize the addressee's whereabouts. A typical postal address is composed from multiple administrative areas, a postal code, a street name, and a house number. As names are reused across different streets often, administrative areas are disambiguating street names. The street name and house number in turn identify a specific house where the post office mail can reach the addressee. Postal codes are the latest innovation in postal addresses. Being invented in the first half of the $20^{th}$ century, postal codes usually encode information redundant to administrative areas. Sometimes postal codes are as accurate as streets or blocks. Besides functioning as a verification to confirm or repair a postal address, postal codes are mainly used to identify a post mail distribution center easily. They thereby enable automated mail sorting. Postal codes are not used by every addressing system. Also the way location data is encoded in postal codes obeys local requirements and varies from country to country. Similar to country specific postal codes, street naming conventions vary by region. While in European countries most streets are named by destinations, areas, landmarks, famous people or events, in USA and Canada many streets are incrementally numbered or labeled and carry directional information with regard to a reference point. House numbering schemes vary just as much as street naming conventions. House numbers may be assigned incrementally or by distance within a street or a block. In the latter case, usually blocks are numbered incrementally. These schemes can be combined with distributing even and odd numbers on different street sides. Some times a single house number is assigned to multiple buildings. A single house number is then not sufficient any more. Additional information identifying a block, a floor or a suite is then added to the postal address. Depending on how important an addressee is in a given region, the location specified in the address may be less specific. Parts of the location may be omitted still assembling a postal address identifying an important addressee unambiguously. E.g., children, addressing post mail to Santa Claus reasonably expect the addressee to be unambiguous, despite the location part of the address being not more specific then North Pole.

Usually computers are only utilized to make sense of location parts of an address. Reaching the addressee (and handing the post mail over) is left to human beings. Because of that there is no unified scheme applied for specifying addressees within large organizations. Arbitrary department hierarchies and case workers may be named as addressees therefore.

Nowadays addresses are present in a variety of sources. Postal addresses are hand-written or printed on letters, forms and packages. Addresses are stored in proprietary schemas in databases of users, customers, orders, etc. GIS databases store addressable entities too, usually with coordinates only not including their postal addresses. Often address parts are also mentioned in free text. From that we can name the following use cases:

Computer systems can be used to

1) derive complete and valid postal addresses from numerical location representations.
2) parse postal addresses from post mail or forms.
3) identify postal address parts in free text documents.
4) construct full and partial valid postal addresses from entities stored in proprietary database schemas.
5) validate, enhance and fix given postal addresses.

 ISBN: 978-1-61208-251-6 

Use case (1) is also referred to as *reverse geocoding*. Measured coordinates, for example, attached to photographs by GPS-enabled cameras, can be presented to users in a more comprehensible way deriving named address parts at the coordinate's location. Use cases (2) and (3) are the daily business at banks, hospitals, government agencies, and such. These organizations do receive filled out forms and informal documents that contain postal addresses to be parsed. Use case (4) is the reverse of (2). It is applied to generate and print an address suitable to post mail from separate address parts. Finally use case (5) is of value for every system gathering postal addresses, to sort out fake or erroneous input. Also having a sparse but unambiguous input address, it is use case (5) to enrich it with all address parts required for a fully specified address.

For dealing with location data contained in postal addresses in an automated way, the common tools are *patterns, gazetteers* and *multidimensional indices*. Patterns allow splitting an address into its address parts and assembling it back according to a predefined format. Gazetteers allow looking up named entities and deriving attached information. The information retrieved from gazetteers may explain hierarchical dependencies between the entities or contain other metadata useful for processing addresses. Finally spatial indices support the reverse look up of geographical entities. Given a point or a polygon, such indices allow deriving address entities that enclose, overlap or touch the specified area.

This paper compares the named tools for processing postal addresses in an automated fashion. Special focus is given to the variety and complexity of postal address formats in use. The goal of this paper is to summarize the merits and demerits of the three tools, thereby identifying gaps for further tooling.

In the following section postal address formats in general are discussed. Next, these tools and their selected variants are analyzed with regards to the defined use cases. Note that it is outside of the scope of this paper to analyze how address texts get digitalized. Hybrid approaches for extending OCR with gazetteers to reach better address recognition while scanning printed or hand written addresses are available [1] but are not included in this analysis.

## II. Postal Addresses

As discussed, there is a variety of postal address formats. As most postal services operate within country borders, most postal address formats differ from country to country. In some countries though, addressing differs within the country depending on the addressee's location being a city, a rural area or a specific region. There are also special address formats for special cases, as the military, where the location of the addressee alters quickly, or shall not be known to the public. Table I illustrates addresses in various formats valid in USA, France, Ireland, and Japan. All addresses

Table I
Example addresses in differing formats within a country.

| | |
|---|---|
| JOHN DOE<br>ABC COMPANY<br>1401 MAIN ST<br>FALLS CHURCH VA 22042-1441<br>UNITED STATES OF AMERICA | SSGT KEVIN BEASLEY<br>UNIT 2050 BOX 4190<br>APO AP 96522-1215<br>UNITED STATES OF AMERICA |
| Monsieur Jean DELHOURME<br>Chez Mireille COPEAU Apartment 2<br>Entrée A Bâtiment Jonquille<br>25 RUE DE L EGLISE<br>CAUDOS<br>33380 MIOS<br>FRANCE | Madame Isabelle RICHARD<br>LE VILLAGE<br>82500 AUTERIVE<br>FRANCE |
| Ms M Sullivan<br>12 Morehampton Road<br>DUBLIN 4<br>IRELAND | Mr J Murphy<br>ABC Company Limited<br>1 Dublin Road<br>Portlaoise<br>CO LAOIS<br>IRELAND |
| Mr Taro Tanaka<br>2-17-10, Aicicho<br>Naka-ku, YOKOHAMA<br>231-0012 JAPAN | Mr Taro Tanaka<br>2338 Shiokawa<br>Maruko-machi<br>Chisagata-gun, NAGANO<br>386-401 JAPAN |

were taken from the address examples provided by the Universal Postal Union UPU[2]. In the first row on the left hand side an typical US address containing *an addressee, house number, street name, town, state, zip-code* (the US-version of a postal code) and *country* is shown. The other address in the same row is addressing military personal, and contains no location information. Only numbers are used that unambiguously identify the addressee but only bear meaning to the Army Post Office of USA. In the next row two French addresses are displayed. While the address on the left refers to an addressee in the city Mios, the address in the right column points to the village Auterive in a rural area. The urban address format also contains the addressee's apartment number and entrance as extended location information. Its parts are *addressee, apartment, entrance, house number, street name, district, postal code, town* and *country*. The rural address format includes *addressee, district, postal code, town* and *country* but names neither a street name nor a house number. In the third row two Irish addresses are confronted with each other. In this case the address formats differ because of regional diversity. Postal codes in Ireland have only been introduced in Dublin. Thus, while the Dublin address contains *addressee, house number, street name, postal code* and *country*, the address with the location information outside of Dublin names the county instead. It consists of *addressee, house number, street name, town, county* and *country*. In the last row two Japanese addresses are shown. The left address points to an urban area. It contains *zone, block, house number, district, town, prefecture, postal code* and *country*, as in Japan street names are not common. Instead houses numbered within blocks which in turn are numbered within zones that assemble a district. If required, a fourth number could be added to that address to specify an apartment. Addressing rural

 ISBN: 978-1-61208-251-6 

areas in Japan uses a different format. The address on the right contains *addressee, house number, district, town, area, prefecture, postal code* and *country*. The area hereby is a part of the prefecture and may contain multiple small towns.

Postal addresses do contain redundant information to make sure that the addressee is identified unambiguously. This redundancy allows postal addresses to be inconsistent. An address may be contradicting containing a valid postal code that does not match the specified administrative areas, or a street name of a street that is located in a district other than the one specified. An address may be incomplete not specifying some vital address parts. Due non-distinct naming of address entities, incomplete addresses may be ambiguous, making post mails undeliverable. Because such addresses do exist, when referring to locations in a computer system, usually numeric location identifiers are used. Such identifiers can be latitude and longitude values of points, a hash value computed from latitude and longitude, or any reference on a Cartesian coordinate system. However, as soon as humans are involved interacting with computer systems about location information, numeric identifiers do no longer suffice. Humans prefer postal addresses mainly for two reasons: First, common adaptation of postal addresses makes them easy comprehensible. Similarly the used scale when referring to time is comprehensible for every human, although it is not based on the decimal system that is employed everywhere else. Second, if an address is not known to a human being, a vague understanding of the location is supported by the named entities of higher hierarchies. In the worst case the only known address part of a given address is the country. Even then, that data is directly derivable data for a human, more then looking at a numeric value.

## III. Evaluation of Postal Address Processing Tools

In this section the tools suitable for processing postal address formats are analyzed with regards to their capability for serving the use cases defined in the introduction. Table II summarizes the outcome. As we will see, multidimensional indices are only capable to look up postal addresses for given coordinates. Patterns support parsing formatted addresses to elementize the addresses parts. Also the reverse procedure assembling an address from separate address parts is executable using patterns. Gazetteers support all the use cases that multidimensional indices and patterns support. That is due to gazetteers containing multidimensional indices and patterns as internal components. In addition, gazetteers may support parsing address parts from unstructured text. As gazetteer contain only valid address parts, they implicitly validate the addresses given.

Table II
Overview of tools and the use cases they serve

| use case | Indices | Patterns | Gazetteers |
|---|---|---|---|
| (1) derive from coordinates | X | | X |
| (2) parse structured | | X | X |
| (3) parse unstructured | | | X |
| (4) construct from parts | | X | X |
| (5) validate | | | X |

### A. Multidimensional Indices

The algorithmic task for efficiently resolving coordinates on a surface into geographical objects covering the coordinates has been solved in a variety of ways. R-Trees [3], Quadtrees [4] and many specialized variants of these [5][6][7][6] make resolving coordinates performant and scalable. A necessary requirement for building such an index is the knowledge about the geographical spread of the postal address parts. With the spatial index at hand it is a straight-forward task to resolve coordinates to the postal addresses that address the same location. Multidimensional indices are therefore a suitable tool for serving use case (1) defined in the introduction.

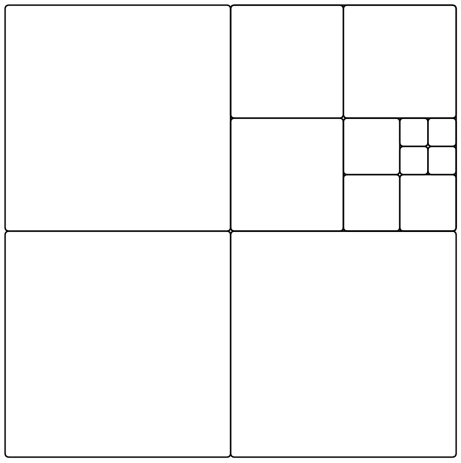


Figure 1. Organizational diagram of a two dimensional Quadtree. Each sector containing more than a specified amount of elements is split in four equally sized sectors.

Figure 1 illustrates the internals of a two dimensional Quadtree. Each square plane is split into four equally sized sectors, as soon as the maximum bound for entries in a single section is exceeded. Depending on the distribution of the entries, some parts of the area are split to smaller chunks, while others remain unsplit. The resulting rectangles are referenced as leaves in an unbalanced tree. The tree grows when on an insert into the Quadtree a sector needs to be split. It then becomes a tree node with four new children.

Often Quadrees are modeled using points with associated location data. The concept of points, e.g., coordinates with zero extent, does not reflect the real world. Houses, blocks, streets, etc. are all objects that have extent. Therefore some Quadtrees are modeled using polygons instead of points. Both points and polygons may be used as input parameters

 ISBN: 978-1-61208-251-6 

to look up address data. Input coordinates may lie exactly between indexed points, or on the border of two polygons. Polygons may contain several indexed points or overlap with multiple indexed polygons to the same part. To resolve these ambiguities, Quadtrees can employ various strategies, i.e., always choose the indexed entity with the lowest coordinate values.

### B. Patterns

Schema patterns are by the simplest tool that can be used to process postal addresses. As patterns are stateless, applying patterns is parallelizable easy. Also there is no data used in a pattern that needs to be kept up to date.

It seems that patterns are a good fit in fulfilling use case (2) from Section I. If the addresses to be parsed contain a delimiter splitting the address parts, simple regular expressions suffice for that purpose. Table III illustrates a Ukrainian sample address with annotated address parts taken from the UPU address samples. The address contains natural delimiters surrounding address parts: Either a whole address part is on its own line, or, if multiple address parts are on the same line, they are separated by commas. Similarly the Japanese address in the fourth row of Table I is machine-readable using simple regular expressions.

That does not apply for every postal address format. The address formats for USA, France and Ireland on Table I use a single space to separate address parts. As some address parts contain spaces in their name, a simple regular expression cannot recognize the exact range of every address part. To approach this problem, Hidden Markov Models (HMM) [8] can be trained to parse the address parts. Conceptually HMMs are statistically learned patterns that pick the most probable path through a finite state automaton to determine the type of input tokens. It is common to use HMMs for elementising address parts as streets names and house numbers from single address lines [9][10].

Both regular expression and HMMs patterns are not flexible enough to support parsing of *all* address formats. Addresses in use contain errors and might have a required delimiter missing. Also, as discussed, not all addresses contain all address parts. Especially addresses of important addressees might omit address parts required by a pattern. Finally as multiple postal address formats are in use, it often is not unambiguous which patter to apply on which address. Therefore Patterns are only serving use case (2) well, if the input addresses to be parsed are of a single known format.

Table III
UKRAINIAN SAMPLE ADDRESS WITH ANNOTATED ADDRESS PARTS

| | |
|---|---|
| Melnik Ivan Petrovitch | addressee identification |
| Vul. Lisova, 2, kv.3 | street, house, apartment |
| s. Ivanovka, Semenivsky r-n, | locality, district |
| TCHERNIGIVSKA OBL. | province |
| 15432 | postcode |
| UKRAINE | country |

Table IV
JOINABLE AND NON-JOINABLE ADDRESS FORMATS

| | | | |
|---|---|---|---|
| Denmark | addressee street house | floor admin2 postcode admin1 | country |
| Germany | addressee street house suite | postcode admin1 | country |
| Denmark joined with Germany | addressee street house suite | floor admin2 postcode admin1 | country |
| Great Britain joined with USA | addressee house street | admin2 admin1 postarea postcode | country |
| Great Britain | addressee house street | admin1 postarea postcode | country |
| USA | addressee house street | admin2 admin1 postcode | country |

Patterns can also be used to process addresses in the opposite direction. Use case (4) requires constructing of valid post addresses from separate postal address parts. Given that the schema of the source containing the address parts is known, applying a pattern rule to assemble these is a very easy step. However, that approach too is constrained by the variety of the postal address formats. If the source contains addresses that have to be assembled according to different postal address formats, the information which pattern to use on which address needs to be available as well.

Figure 2 shows how an address is assembled using a pattern. The separate address parts are retrieved from a proprietary database schema. The pattern instructs in which order to assemble the address parts and what separators to use. Note that not all the address parts retrieved from the database are used.

The UPU follows this approach with their product ”International Postal Address Components and Templates” (UPU S-42). It specifies rules for assembling the 35 identified address parts. UPU S-42 covers 246 countries but it needs to define regional postal address formats too. To estimate the amount of schemes in UPU S-42, for a reduced address part set of 17 address parts, address samples provided by the UPU have been annotated. The resulting 450 address formats have been joined as shown in Table IV: Two address formats may be joined only if the address parts present in both address sets are ordered equally. They are then combined into a joined address format that contains all address parts from both address formats, respecting the order of address parts in both original formats. In Table IV the two combined address formats may not be joined, because the orders of street name and house number and administrative level one and postcode are conflicting. The joinable address formats have been joined so that the amount of address formats was reduced to 41. The same effort has been undertaken

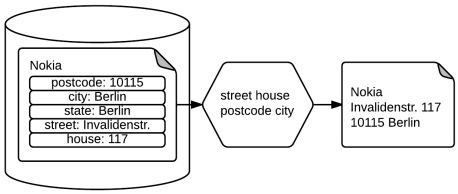



Figure 2. The address of the Nokia headquarters Berlin assembled by a pattern from a proprietary data base schema.

 ISBN: 978-1-61208-251-6 

by the UPU. UPU S-42 is a ready alternative for composing valid postal addresses without the upfront analysis. However, it requires the data to be as fine-grained as the scheme used in UPU S-42 using the same 35 parts of a postal address. For denormalized data bases, and if not all address formats identified by UPU need to be constructed, a light-weight approach using patterns suffices, to construct postal addresses from their parts.

### C. Gazetteers

As multidimensional indices discussed in Section III-A, gazetteers index location parts entities. Unlike multidimensional indices gazetteers index names of entities instead of their position. This way gazetteers support querying for address parts by name. Retrieved entities usually have attached metadata describing relationships to other named entities. This spanned structure may be browsed.

Gazetteers can be used for detecting address parts in free text and in structured addresses. Those are use cases (2) and (3) from Section I. For both use cases, as no clear boundaries of a single address part is described. It is not obvious what part of an address or a free text to use when querying a gazetteer service. Some gazetteers may be able to retrieve *all* address parts it finds in a query, solving the splitting problem transparently to the user. Others require multiple queries of *n-grams*. Splitting a unstructured text or an address of an arbitrary format is complex in implementation and execution. Therefore most gazetteers are not independent of the input but define a fixed postal address format that they support.

With all address parts being indexed in gazetteers, only valid address parts are retrievable. Using the relationship data – if present in the gazetteer – the parsed addresses are implicitly validated. That serves use case (5) from the introduction. Often the relationships of location entities are represented via their numerical location and spread. E.g., a gazetteer could derive that New Jersey is a part of the USA knowing the geographical spread of both entities. That would be implemented using a multidimensional index as discussed in Section III-A, enabling gazetteers serving use case (1) as well.

Named location entities are the base units for gazetteers. As entity names alone are ambiguous, gazetteers have to qualify found entities using other address parts. Particularly, to distinguish between equally named location entities, gazetteers have to name all location entities enclosing the equally named ones. This happens according to a schema from which, in combination with a pattern, an address can be constructed. Many gazetteers implement that feature on their end, solving use case (4) from Section I.

A gazetteer protocol has been defined by the Open Geospatial Consortium (OGC) [11]. The OGC Best Practices Document defines a gazetteer as ”a database used to translate between different representations of geospatial references, such as place names and geographic coordinates” (Section 4, page 12). The OGC gazetteer does not define a specific address format for processing, but defines a hierarchical model of named location entities. These entities are associated to features, which are not necessarily spatial. This way a gazetteer that complies with the OGC protocol supports looking up location entities by name or coordinates. Queries to the gazetteer are allowed to specify a filter on non-spatial features, which reduce the result list. Optionally retrieving location entities by a gazetteer specific ID is supported too. The protocol allows browsing the relationship of entities directly via links between them, or indirectly by fetching all entities that are within a certain bounding box. A gazetteer implementing the OGC protocol and filled with location parts of addresses would support use cases (1) and (4) only if a set of patterns matching the gazetteers internal schema was provided. The gazetteer is incapable of detecting multiple named entities in a query. It therefore requires the client to slice the input according to the boundaries of address parts, to support use cases (2) and (3). As any gazetteer, it does fulfill use case (5) as only valid location address parts are identified. Browsing the links between location parts allows discovering contradicting addresses.

Another example of a gazetteer is the Google Geocoder [12]. Unlike the multi-purpose protocol of OGC, Google’s gazetteer focuses on translating addresses into latitude and longitude coordinates and vice versa. For that the gazetteer protocol accepts requests with addresses only. No specific address format is required. As the OGC gazetteer, Google’s gazetteer support filtering results by specifying selected address parts as required. The address parts *route, locality, administrative area, postal code* and *country* are supported for filtering. Apparently Google’s gazetteer internally defines the patterns required to construct an address, as an assembled address is always part of each result. Google’s gazetteer supports use cases (1), (2), (4) and (5) as defined in the introduction. As the OGC gazetteer, for detecting address parts in free text (use case (3)) it remains to the client to slice the text into chunks that contain single address parts.

Many other implementations of gazetteers exist. For

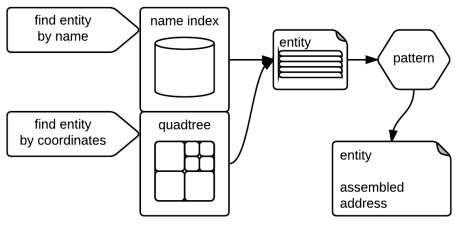



Figure 3. Gazetteer Data Flow. Depending on the request containing a name or a coordinate, either the name index, or the multidimensional index is queried. Addresses of retrieved entities are constructed through a pattern.

 ISBN: 978-1-61208-251-6 

example, Densham et al. [13] describe a system that is extending a gazetteer with capabilities to parse free text. Clough et al. [14] have built and measured a system that assigns coordinates to web documents.

Overall gazetteers are more complex than patterns or multidimensional indices, as most gazetteers contain both patterns and multidimensional indexes internally. That implies that gazetteers have to maintain a list of patterns to construct addresses according to various formats. Also as multidimensional indices, the data on location entities needs to be kept up to date. Finally indexing named entities is a non-trivial task too. In sum, that makes gazetteers a tool that is complex to manage, while it does support more use cases simultaneously.

In Figure 3 a basic gazetteer set up is visualized. It includes a text index for name queries, a multidimensional index for coordinate queries and a pattern to assemble address parts of results. The visualized gazetteer stores entities with their address parts in both indices. To save space, other gazetteer implementations might only store references in their indices that point to a common data source. The entities are stored in a schema that can be assembled to correct addresses using patterns.

## IV. Conclusion

The paper evaluated *gazetteers, multidimensional indices* and *patterns* as tools for processing postal addresses. It showed that there is no general pattern that is suitable to support parsing or constructing addresses, because multiple address formats need to be supported. The option to use specialized patterns for individual address formats is not available for all address sources, depending on the extent of regions that are being covered. Multidimensional indices resolve numeric location references as coordinate pairs into location entities. These however still need to be assembled to valid addresses. Also multidimensional indices need to be kept up to date for resolving addresses accurately. Gazetteers are the most versatile tool for processing postal addresses. Depending on implementation and interface, gazetteers can support parsing structured addresses and free text, constructing and validating addresses and resolving addresses from numerical references. To achieve that, gazetteers contain patterns and multidimensional indices internally, in addition to the text index for looking up named entities. Because gazetteers are such a complex tool there are efforts joining gazetteers with other address related tools to achieve better performance.

There is a gap in complexity between gazetteers and patterns. While gazetteers serve more use cases they also are by magnitudes more complex compared to simple address format patterns. One option to close this gap is to split address patterns in logical units of multiple address parts. These sub-patterns could be used in gazetteers that support retrieving combined address parts if these belong together. Another method to combine these tools are smart patterns that rely on data to automatically determine how to assemble a complete and correct address, depending on country and region.

## References

[1] S. Srihari, "Recognition of handwritten and machine-printed text for postal address interpretation," *Pattern recognition letters*, vol. 14, no. 4, pp. 291–302, 1993.

[2] "Universal Postal Union," http://www.upu.int, Jun. 2012.

[3] A. Guttman, "R-trees: a dynamic index structure for spatial searching," vol. 14, no. 2, 1984.

[4] H. Samet, "The quadtree and related hierarchical data structures," *ACM Computing Surveys (CSUR)*, vol. 16, no. 2, pp. 187–260, 1984.

[5] R. Kothuri, S. Ravada, and D. Abugov, "Quadtree and R-tree indexes in oracle spatial: a comparison using gis data," in *Proceedings of the 2002 ACM SIGMOD international conference on Management of data*. ACM, 2002, pp. 546–557.

[6] N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger, *The R*-tree: an efficient and robust access method for points and rectangles*. ACM, 1990, vol. 19, no. 2.

[7] Y. Kim and J. Patel, "Performance comparison of the R*-tree and the quadtree for knn and distance join queries," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, no. 7, pp. 1014–1027, 2010.

[8] L. Rabiner and B. Juang, "An introduction to hidden markov models," *ASSP Magazine, IEEE*, vol. 3, no. 1, pp. 4–16, 1986.

[9] A. Kornai, "An experimental hmm-based postal ocr system," in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 4. IEEE, 1997, pp. 3177–3180.

[10] V. Borkar, K. Deshmukh, and S. Sarawagi, "Automatically extracting structure from free text addresses," *IEEE Data Engineering Bulletin*, vol. 23, no. 4, pp. 27–32, 2000.

[11] J. Fitzke and R. Atkinson, "OGC best practices document: Gazetteer service-application profile of the web feature service implementation specification-0.9. 3," *Open Geospatial Consortium*, 2006.

[12] "The Google Geocoding API," https://developers.google.com/maps/documentation/geocoding, Oct. 2012.

[13] I. Densham and J. Reid, "A geo-coding service encompassing a geo-parsing tool and integrated digital gazetteer service," in *Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references-Volume 1*. Association for Computational Linguistics, 2003, pp. 79–80.

[14] P. Clough, "Extracting metadata for spatially-aware information retrieval on the internet," in *Proceedings of the 2005 workshop on Geographic information retrieval*. ACM, 2005, pp. 25–30.

 ISBN: 978-1-61208-251-6 

# TerraHidro: A Distributed Hydrology Modelling System With High Quality Drainage Extraction

*Sergio Rosim, João Ricardo de Freitas Oliveira, Alexandre Copertino Jardim, Laércio Massaru Namikawa, Camilo Daleles Rennó*
Image Processing Division
*National Institute for Space Research*
São José dos Campos, Brazil
{sergio, joao, camilo, laercio, alexandre}@dpi.inpe.br

***Abstract*—This paper presents the current development stage of TerraHidro, a Distributed Hydrology Modeling System created to design Geographic Information Systems (GIS) applications for water flow in hydrographical basins. TerraHidro proposes a different computational representation to deal with water flow in GIS applications. At first, this paper presents the conceptual model of TerraHidro, where the structure used to develop applications is independent from the structures used to extract local flow paths. Since the local flows are primary information for network drainage extractions, the Priority First Search (PFS) method used by TerraHidro to extract these local flows is presented here. PFS method also gives realistic results in flat areas and eliminates all spurious pits in the DEM data with small modifications in the elevation values. Examples of drainage networks are presented and a comparison between TerraHidro and ArcGIS Hydro Tools are presented.**

***Keywords; local flow; distributed hydrological models; PFS method; drainage network***.

## I. INTRODUCTION

The local flow distribution in a water basin is the most important factor in order to model distributed hydrological systems that aim at hydrological resources management. The underlying premise is that terrain topography is the main landscape contributor in defining these local flows [1][2]. Mathematical representations of terrain topography and sets of functions to extract superficial local water flow have been explored by GIS users for a long time. The basis for terrain topography representation in GIS is to partition the region extents. A *Cell* is the unit of this partition set and the *local flow* is the *water flow* for each *cell* considering the status of its neighbor *cells* according to a specific chosen neighborhood rule.

The most common data structures found in GIS libraries and systems for terrain representation dedicated to hydrological modeling are the DEM (Digital Elevation Model - DEM) [3] with regular grids, Irregular Triangular Networks – TIN [4], Contour Lines based representation [5], and Irregular Polygons Tessellations [6]. Each chosen surface representation carries its own local water flow extraction functions and its own local flow data structure. The local flow representation is dependent of the data structure used to represent the terrain topography. For instance, DEM local flow extraction uses the 8-neighbor concept and creates a local flow representation, called *Local Drain Direction* (LDD) [3]. It is stored in the same structure DEM, forming a regular grid of local flow directions.

This situation takes hydrological modeling to a condition where there is a strong coupling between the quality of the local flow representation and the parameters for the used terrain data structure, in particular its spatial resolution with direct implications on the model outcomes [7]. Distributed hydrological modeling environments normally assume a unique data structure for terrain representation. This may simplify software development but it can't make use of the properties of the other terrain data structures. The concept proposed by the TerraHidro [8][9][10][11], a distributed hydrological system, assists the simplification of software development and at the same time allows the use of different terrain data structures. In this way, the decoupling of terrain data structure and local flow data structure eliminates the need to code a given operator for each one of the used terrain structures [8].

Currently, TerraHidro works only with DEM structure to extract and develop applications and it is a plugin of the geographic visualizer TerraView that loads and stores data in a geographical library called TerraLib [2], an open source geographical library implemented in C++ language that has been developing at the Image Processing Division of National Institute for Space Research, in São José dos Campos a Brazil city situated in the São Paulo region. This approach has allowed TerraHidro project team of designers and programmers keep focused on the development of system functionality of TerraHidro. TerraView main goal is to make available to the GIS Community an easy geographic data viewer with resources that include database queries and data analysis, exemplifying the use of the TerraLib library.

 ISBN: 978-1-61208-251-6 

TerraLib is an open-source GIS software library. TerraLib supports coding of geographical applications using spatial databases, and stores data in different DBMS including MySQL, PostgreSQL and other databases. Figure 1 shows relationship among Terrahidro, TerraView and TerraLib.

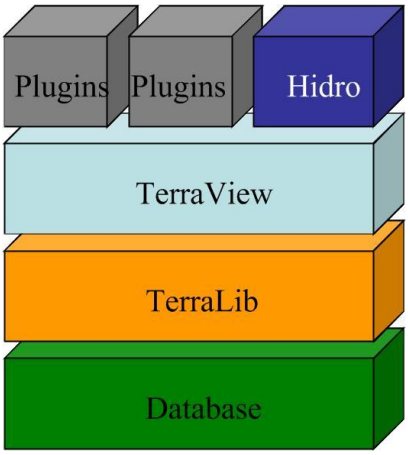



Figure 1 TerraHidro, TerraView and TerraLib relationship

The focus of TerraHidro development has been to obtain the best quality for each generated result. Optimizations, such as reducing processing time, are considered whenever they are not detrimental to the results quality.

This paper is structured as follows: Section two describes the steps required to obtain the drainage network, Section three presents the method used by TerraHidro to define the local flows, Section four shows the results obtained, and Section five has the conclusions.

## II. DRAINAGE EXTRACTION AND BASIN DELIMITATION

Drainage has been used for several applications involving water resources. To extract the drainage, first of all, the local flow and the accumulation area must be determined. Local flow is a flow between two neighbor cells considering the steepest downstream path among a cell and each of its eight neighbor cells. Figure 2 shows this concept.

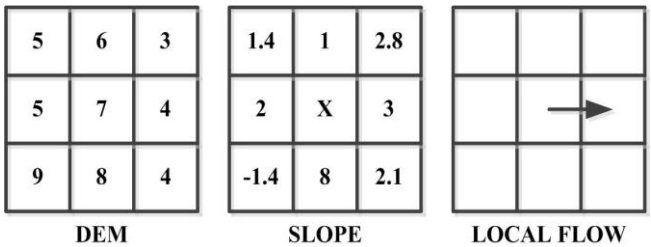



Figure 2 Local flow extraction to X cell

After this step, the accumulation areas are calculated from the local flows. Each Y cell receives a value that is the size of the area of all cells that are on the path arriving at Y cell. The next step requires the definition of the accumulation area subset called drainage network. The user defines a threshold value and all grid cells with values equal or greater than this threshold value are defined as drainage cells. At this point, TerraHidro can define the river reaches that delineate the drainage segments. The segments are between the water sources and junctions, between junctions, and between junctions and drainage mouth. Basin delimitation, the next processing step, can be executed by selecting one or more points on the drainage. TerraHidro finds the basin for each given point or for each river reach. In the end, the basins can be used as grid cell, as vector forms. Figures 3, and 4 present this concept. Figure 3 has drainage network segments. Each segment is represented by a different color. Figure 4 shows the watershed for each segment.

## III. PFS AND OTHER DRAINAGE EXTRACTION METHODS

The extracted drainage is the basic information to develop hydrological applications. For this reason, the method used to extract the drainage must minimize the effects of errors from DEM generation, such as the ones present in the Shuttle Radar Topographic Mission (SRTM) data set with 90 meters of horizontal resolution is used in this work. Although the SRTM data is used worldwide, it contains a large number of spurious pits among other problems. Spurious pits are false end points of flow. The drainage extraction should ensure that the drainage is connected in the adopted resolution with small changes in the original SRTM values, generating the most real drainage possible.

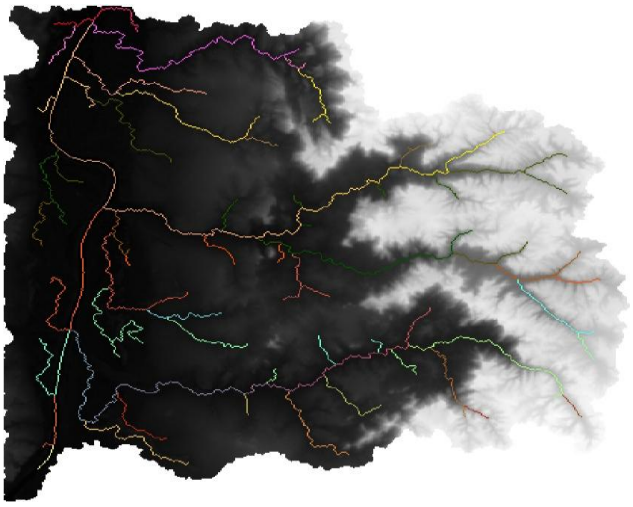


Figure 3 River segments.

Given that the core of water resource applications are drainage networks, the drainage extraction method used by TerraHidro will be described in the next section.

 ISBN: 978-1-61208-251-6 

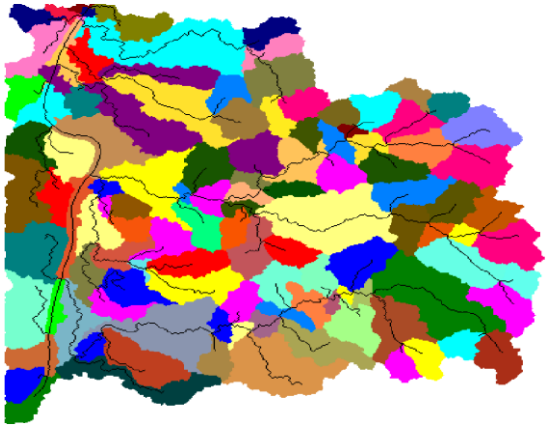


Figure 4 Watersheds delimitation. Each watershed corresponds one river segment.

TerraHidro uses two approaches to solve the spurious pits problem. In the first one, the mean is calculated for each pit using the neighborhood 8 of the pit. If the pit remains, or generates a new pit in one of the neighbors, their value is not replaced by the mean, and then it uses the PFS method to solve the problem [12]. The basic idea of this method is to link a pit with a nearby grid cell with lower elevation creating a optimum path between both cells. Next step is to define a new elevation value for each grid cell in this path creating a monotonic down slope from pit cell to lower elevation cell. By applying the method for every pit, PFS creates a fully connected drainage network generating elevation value changes only in the path. Figure 4 shows the diagram of TerraHidro method to eliminate pits. PFS has presented better results when compared with others methods according studies made in [13]. Figure 5 shows the diagram of TerraHidro method to eliminate pits.

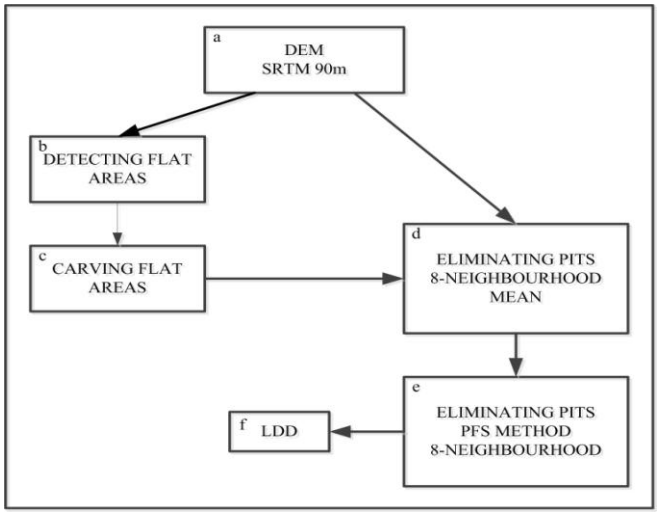



Figure 5 Pit removal schema.

The pit removal schema follows the sequence presented below:

(a) In the first step, the SRTM data set is inserted into the TerraHidro system. Next, the user chooses between two paths: either create drainage in the flat regions and then proceed to the pit elimination or proceed directly to eliminate the pits.

(b) SRTM creates spurious flat areas on large water bodies. TerraHidro recognizes and delimits these areas to be used in the carving processes. This step is necessary because PFS method don't produce good results in the flat areas [14].

(c) The carving process creates a down path slope from the border of the flat area until reaching the center of this area. This is executed for every point on the border of the flat area. The drainage flows at the center of the flat area. Figure 6 shows the schematic representation of the carving process and Figure 7 presents two real geographical areas in the Brazilian Amazon region. The water body areas were highlighted with yellow ellipses to exemplify this type of case.

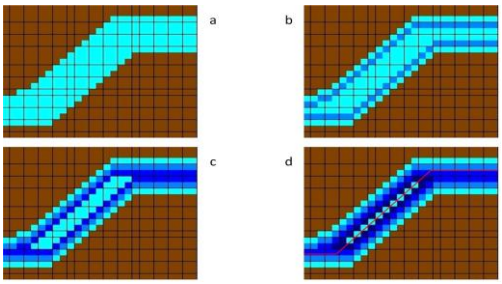



Figure 6 (a) Plane area delimited; (b), (c) intermediate carving processes; (d) water body carved with water flow path (red line).

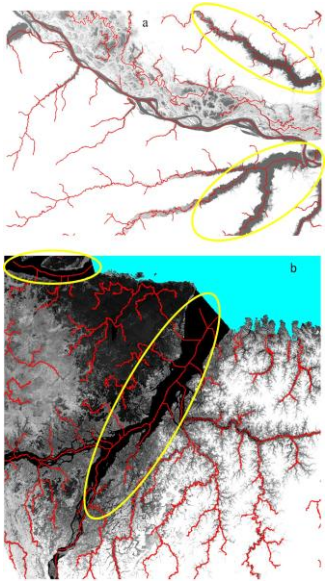



Figure 7 Drainage networks of plane areas (yellow ellipses) (a) Purus basin and (b), (c) Tapajós basin.

(d) For each grid ***C*** cell of altimetry that contains a pit, TerraHidro calculates the mean values of its eight

neighboring cells. If this procedure does not create a new pit in this neighborhood, the value computed is assigned to ***C*** cell, thus eliminating the pit. Otherwise, the pit of the ***C*** cell remains. Figure 8 presents an example of this pit removal procedure.

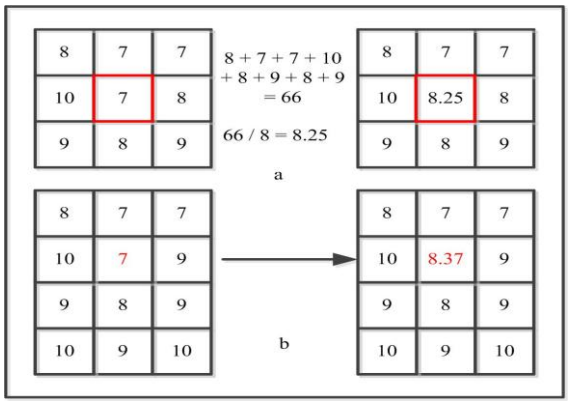



Figure 8 Pit removal by 8-neighbourhood media average calculus. (a) Without generating new pit; (b) generating new pit..

(e) Pits that are not removed by the mean calculation procedure are eliminated using Priority First Search - PFS method. This method finds a path between a pit and a nearby grid cell with lower elevation. For each neighboring cell of the one containing the initial pit, the differences between elevation values are calculated and stored in a queue. The path to the smallest value is selected. When there are equal differences for more than one neighbor, the cell that is closest to the one that contains the pit will be selected. The process is repeated from the selected cell. The minimum distance is always calculated in relation to the position of the cell containing the initial pit when points differences are tied. At this point, the new elevation value for each cell belonging to the path between the two pits is calculated accordingly to its distance from these pits. Figure 9 shows this procedure.

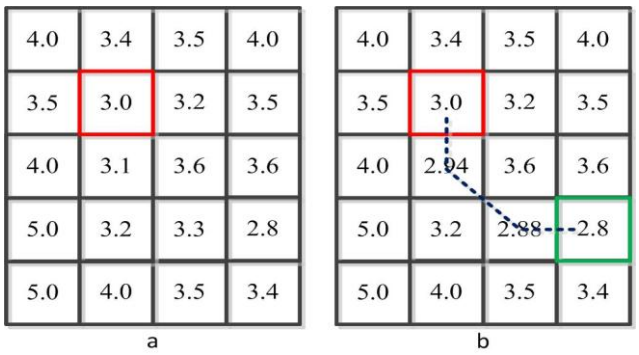



Figure 9 Carving processes to eliminate pits. (a) Initial pit identification (red color); (b) path found by PFS method (final pit in green color).

The modified PFS method eliminates all the pits in the existing elevation grid ensuring flow inside the entire study region.

## IV. RESULTS

TerraHidro has been used in large geographical areas. The focus has been to test TerraHidro in actual extensive watershed areas using computers with typical RAM memory capacity, for example, 3 GB. TerraHidro has extracted drainage networks successfully for the test areas eliminating all pits. This qualifies TerraHidro as a robust system, giving also very good results in terms of drainage network extraction and watershed delimitation. Amazonian and sub Amazonian basins and South America region have been used by TerraHidro to extract drainage networks, using the SRTM data set with 90 meters spatial resolution.

Figure 10 shows the drainage network for the Taquaruçu River basin region. The DEM of this region has 2.024 rows, 1.875 columns (3.734.280 image size) and 396.769 pits.

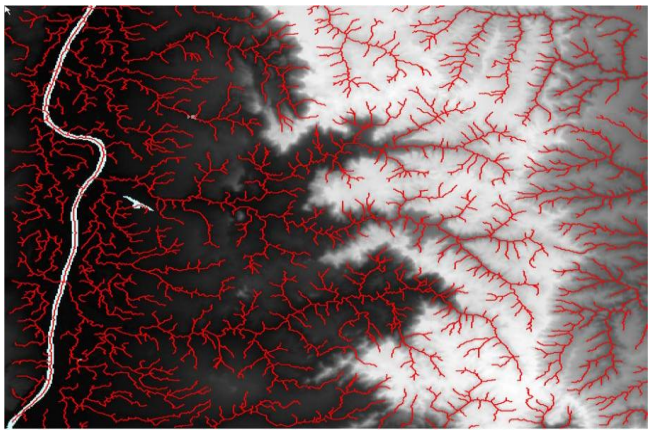


Figure 10 Drainage network of Taquaruçu basin (red color)green).

Figure 11 Presents results of the drainage extraction of Xingu River region, another Amazon sub basin. The information of this DEM is: 15.962 rows, 7.202 columns (144.958.324 image size)) and 6.472.113 pits. The image on the right side is the zoomed area in of the yellow rectangle in the image on the left side.

The next watershed is the Tapajós River Amazon sub basin, which can be seen in Figure 12. This region has 19.201 rows, 9.601 columns (184.348.801 image size) and 8.647.984 pits. In the image on the right side, it is possible to see the carving in the plane area into the river (up to down in the middle of image).

Figure 13 presents Purus River, that is another Amazonian sub basin. Also here it is possible to see the drainage flowing by the middle of the large rivers correcting these plane areas. The data of the Purus River basin are: 12.000 rows, 15.600 columns (187.2000.000 cells) and 13.279.394 pits.

 ISBN: 978-1-61208-251-6 

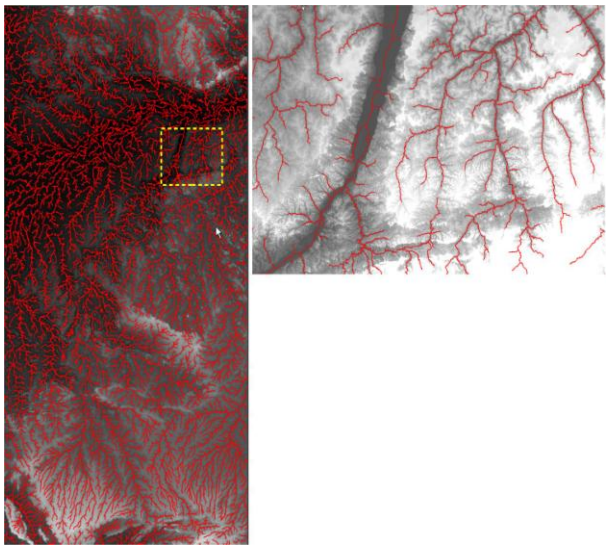


Figure 11 Drainage network of Xingu basin.

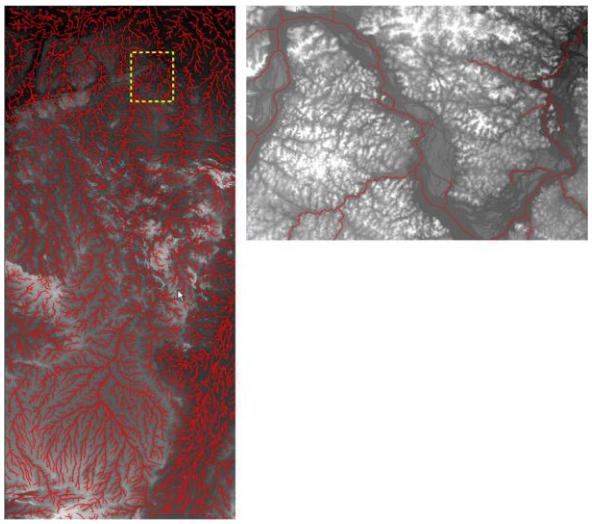


Figure 12 Drainage network of Tapajós basin.

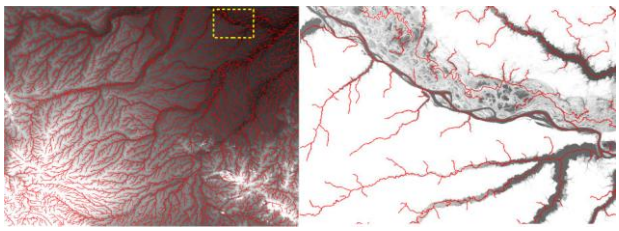


Figure 13 Drainage network of Purus basin.

The last Amazon sub basin processed by TerraHidro was the Tocantins River basin. Figure 14 shows the drainage extracted by TerraHidro from a region with 21.602 rows, 14.402 columns (311.112.004 image size)) and 15.893.139 pits.

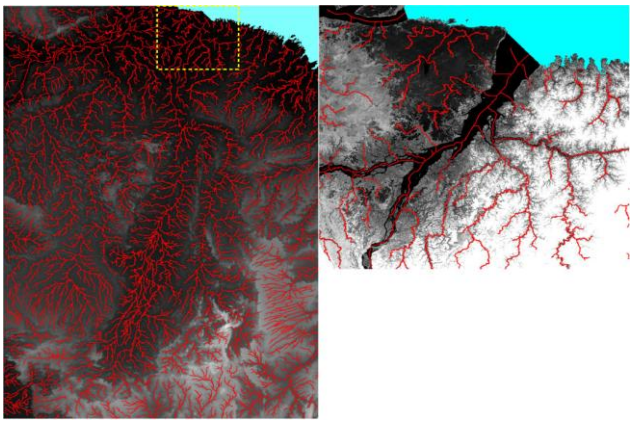


Figure 14 Drainage network of Tocantins basin.

The whole Amazon basin was also processed. Figure 15 shows in red color the delimitation of this basin. In green color, it is possible to see part of the Amazon basin river network. The data of the Purus River basin are: 32,400 rows, 38,400 columns (1,244,160,000) and 65,670,466 pits.

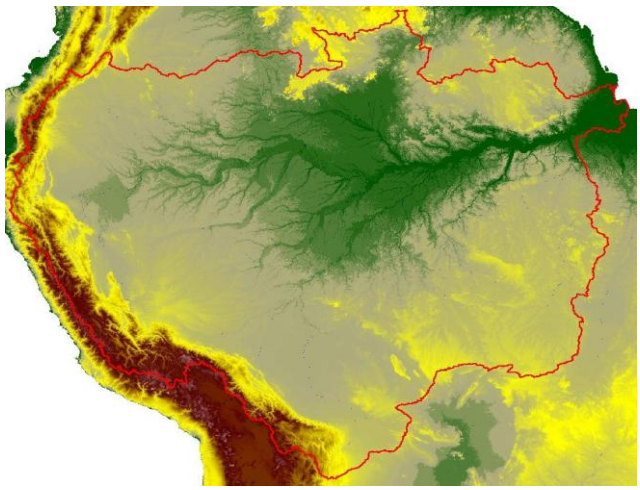


Figure 15 Drainage network of Amazonian basin.

The whole South America region was also processed. Figure 16 presents the drainage networks with the most important South American rivers. The information of the DEM are: 60.001 rows, 84,001 columns (5,040,144,001 image size) and 161.135.443 pits.

 ISBN: 978-1-61208-251-6

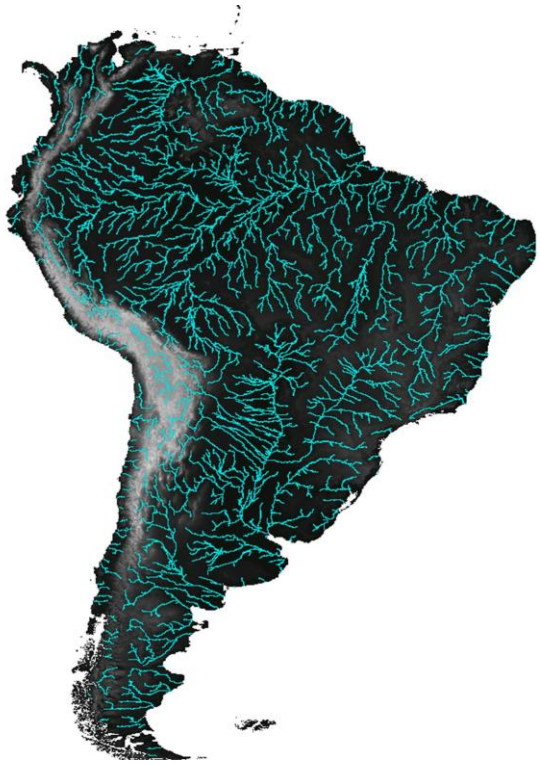


Figure 16 Drainage network of South America region.

Drainage networks have also been extracted for the African countries of Somalia, Kenya and South Sudan. Figures 17, 18 and 19 show the drainage (blue color) and the corresponding watersheds for each drainage segment.

Figure 17 Drainage network and watersheds for each drainage segment for Somalia country.

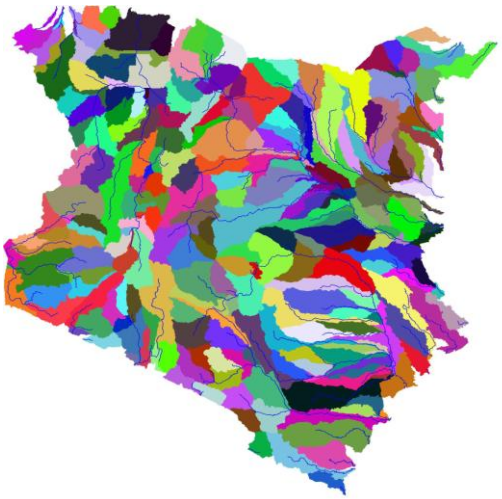


Figure 18 Drainage network and watersheds for each drainage segment for Kenya country.

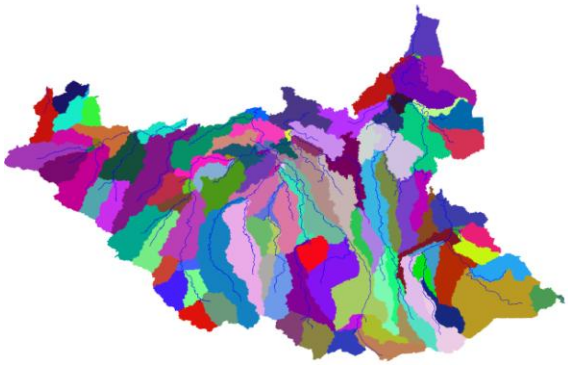


Figure 19 Drainage network and watersheds for each drainage segment for South Sudan country.

Finally, a comparison between ArcGIS Hydro Tools and TerraHidro regarding drainage network quality is presented in the Figure 20. In the top figure, the ArcGIS Hydro Tools results appear in blue lines and the TerraHidro ones are in red lines. The results obtained by ArcGIS in the flat areas indicates that the method D8 is used directly in these areas without prior corrections. The parallel lines generated by ArcGIS indicate incorrect representation of the drainage within the flat areas. TerraHidro, for the same areas, identify automatically each flat areas creating a path through the middle of the flat area, following its longitudinal direction. All path into flat area arrives at this longitudinal path in a physical way.

This comparison aims to show that TerraHidro have a method to determine the path in the flat areas generating coherent drainage networks. Parallel lines that appear in ArcGIS Hydro Tools drainage networks are caused by the method used in ArcGIS that converts pits in flat areas and, after that, uses D8 method in these flat areas. The method used by TerraHidro system does not produce these errors

 ISBN: 978-1-61208-251-6 

because a path between from a pit is found without the need to create flat areas.

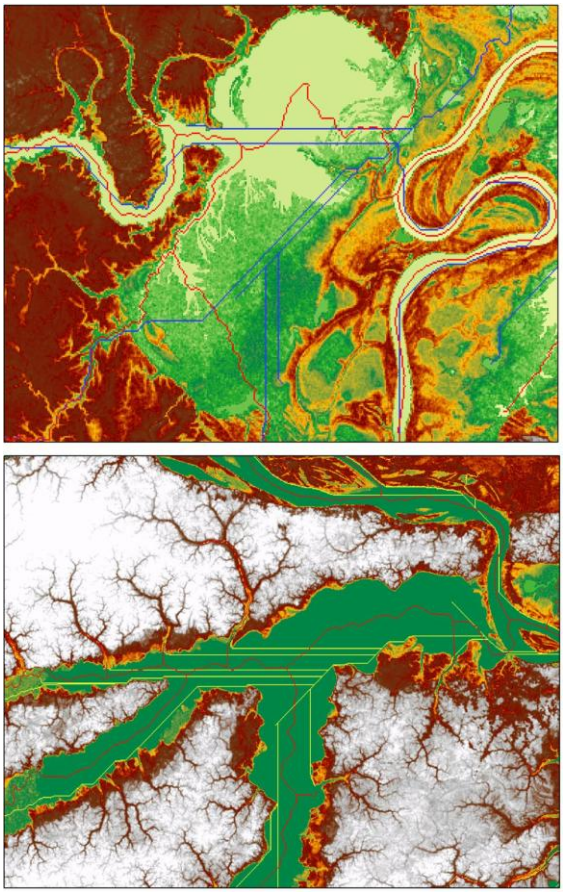


Figure 20 Comparison of quality of drainage network extraction between TerraHidro and ArcGIS Hydro Tools.

## V. CONCLUSION

The concept of the TerraHidro distributed hydrology modeling system and its current development stage were described here. The PFS method to extract drainage network was shown together with improvements to define drainage into the flat areas such as large rivers and lakes. In addition, the 8-neighbor mean to eliminate pits without creating others in its neighborhood was presented.

Several results were presented here. All of them used large regions to show the quality of the extracted drainage and to test the robustness of TerraHidro. Drainage extractions were executed using a PC computer with 2 GB of RAM memory.

A comparison of the drainage extracted by TerraHidro and ArcGIS Hydro Tools was made for Purus Amazon sub basin. The comparison showed that TerraHidro produces more realistic drainage than ArcGIS Hydro Tools do.

All processing used SRTM 90 meters data set. Other data sets such as ASTER-GDEM will be used in the future.

## ACKNOWLEDGMENT

This article is part of Automatic Protection Delineation Brazilian Areas project entirely supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP a Brazilian government Agency that finances studies and projects in the area of science and technology.

## REFERENCES

[1] C. D. Rennó, "Construção de um Sistema de Análise e Simulação Hidrológica: Aplicação a Bacias Hidrográficas (in portughese)," Ph.D. thesis, National Institute for Space Research, São José dos Campos, 2003.

[2] J. F. O'Callaghan and D. M. Mark, "The Extraction of Drainage Networks from Digital Elevation Data," Computer Vision, Graphics, and Image Processing, 28:323–344, 1984.

[3] P. A. Burrough and R. A. "McDonnell, Principles of Geographical Information Systems." Oxford University Press, New York, 1998.

[4] L. P. Chew, "Constrained Delaunay triangulations," Algorithmica 4(1):98-108, 1989.

[5] W. R. Dawes and D. L, Short, "TOPOG Series Topographic Analysis and Catchment Drainage Modelling Package: User Manual," Canberra, CSICO, 1988.

[6] J. Fairfield and P. Leymarie, "Drainage Networks from Grid Digital Elevation Models" Water Resource Research, 30(6):1681–1692, 1962.

[7] J. Garbrecht, F. L. Ogden, P. A. DeBarry and D. R. Maidment, "GIS and Distributed Watershed Models. I:Data Coverages and Sources," Journal of Hydrologic Engineering, 6(6):506-514, 2001.

[8] S. Rosim, "Estrutura baseada em grafos para representação unificada de fluxos locais para modelagem hidrológica distribuída. 2008. 110 p. (INPE-15320-TDI/1363). (PhD thesis in applied computation) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2008.

[9] S. Rosim, A. M. V. Monteiro, C. D. Rennó, J. R. F Oliveira, "Terrahydro - A distributed hydrological system using graph structure for unified water flow representation," Proc. of IEEE Geoscience and Remote Sensing Symposium (IGARSS), Vancouver, 2011.

[10] E. S. Abreu, S. Rosim, C. D. Rennó, J. R. F. Oliveira, A. C. Jardim, J. O. Ortiz, L. V. Dutra, "Terrahidro a distributed hydrological system to delimit large basins," Proc. of IEEE Geoscience and Remote Sensing Symposium (IGARSS), Munich, 2012. v. 1, pp. 300-301.

[11] S. Rosim, A. M. V. Monteiro, C. D. Rennó, J. R. F. Oliveira, "Uma ferramenta open source que unifica representações de fluxo local para apoio à gestão de recursos hídricos no Brasil," Informática Pública, v. n. 1, p. 29-49, 2008.

[12] A R. Jones, "Algorithms for using a DEM for mapping catchment areas of stream sediment samples", Computers & Geosciences 28 (2002) 1051–1060.

[13] B W. Collischonn, D. C. Buarque, A. R. da Paz, C. A. Bulhões Mendes, and F. M. Fan, "Impact of Pit Removal Methods on DEM Derived Drainage Lines in Flat Regions", AWRA 2010 Spring Speciality Conference Orlando, FL.

[14] R. Jana, T. V. Reshmidevi, P. S. Arun, T. I. Eldho, "An enhanced technique in construction of the discrete drainage network from low resolution spatial database," Cumputers & Geosciences archive, v. 33, I. 6, June 2007, Pergamon Press, Inc. Tarrytown, NY, pp. 717-727, 2007.

 ISBN: 978-1-61208-251-6 

# Indoor-space Awareness using Descriptive data

Kang-jae Lee
Department of Geoinformatics
University of Seoul
Seoul, South Korea
kjlee@uos.ac.kr

Hye-Young Kang
Department of Geoinformatics
University of Seoul
Seoul, South Korea
hyezero@gmail.com

Jiyeong Lee
Department of Geoinformatics
University of Seoul
Seoul, South Korea
jlee@uos.ac.kr

*Abstract*— **Wide use of locational information has recently contributed to many GIS application services in a variety of ways, for example, shopping, tourism and transportation. In particular, these human behaviors have been widened from outdoor into indoor-space due to the appearance of large scale and complex buildings. Thus, indoor localization for various location based services has been researched in indoor-space using wireless LAN, RFID or Bluetooth instead of using GPS. However, these positioning technologies require intensive computation and have a limitation to retain stable accuracy. Thus, in this paper, we introduce Indoor-space Awareness Architecture using descriptive data such as names or phone numbers from images captured from mobile devices. OCR engine is especially applied in the architecture to recognize text from the descriptive data.**

***Keywords - Indoor-space Awareness; Architecture; Descriptive data; Optical Character Recognition; Geocoding.***

## I. INTRODUCTION

The use of locational information has recently increased in web map services and map applications with a variety of information. As spatial information, the locational information is a basically used with non-spatial information for providing users with useful information such as shopping, tourism and transportation information near particular locations on the web map. Furthermore, as smart phones come into wide use, location-based applications are being developed steadily due to built-in GPS. So the locational information is used in real time received from GPS satellites. As a kind of positioning technology, GPS is the most useful system to obtain locational information for outdoor environments. However, in the case of indoor space, the GPS has a limitation on the accuracy of positioning since GPS signals cannot be directly come into buildings without any interference on the signals.

For the indoor positioning, several techniques are applied such as Wireless LAN [11, 12], RFID [4], Bluetooth [8] and InfraRed [3]. In spite of the application of these wireless techniques, awareness of location in indoor-space is restricted in the case of no infrastructure such as access point and sensors. Also, they have variable localization accuracy due to walls and columns. Thus, the alternative methods for the indoor localization need to be developed. In order to meet the minimum requirements, the alternative methods have to be simple to estimate locations and obtain locational data using information in indoor-space. So we designed Indoor-space Awareness Architecture using descriptive data which is readily available from stores, office rooms and convention halls in indoor space. So modified geocoding method is applied for matching descriptive data with locational data. In addition, for recognition and extraction of characters or words from the descriptive data, Optical Character Recognition (OCR) is applied. The OCR technology has grown with template matching and structure analysis in the pattern recognition field [13]. The descriptive data includes store names of complex shopping malls, room numbers of hotels or offices and phone numbers for deriving locational information in indoor-space.

Implementation of this research is conducted on Android OS for smart phones. Smart phones are widely used and most of them have camera modules which are appropriate for capturing images simply and quickly for acquiring descriptive data. As a result, an application based on Android OS is developed according to the structure of Indoor-space Awareness Architecture. Furthermore, a query service based on the architecture is implemented using related search terms.

## II. RELATED WORKS

Numerous researches have been carried out regarding location awareness and navigation using a variety of data and technologies. For this paper, the main issue is how image data is handled with locational information using geocoding, thus the previous researches on image-based positioning and geocoding methods are reviewed.

Image-based location awareness has been developed with geo-tagged image data. The research on the algorithm for estimating geographic information from image data was conducted using a purely data-driven scene matching approach [9]. According to the research, over 6 million GPS-tagged image data was stored in database and some of them were partially used for the evaluation of the method. Images for the test were matched with database in accordance with color patterns and texture features so as to recognize correlation of between the images and geographic locations. And the geo-locations of the images were estimated using a set of GPS tagged nearest neighbors. Another research on the system for managing digital photographs with geographic location tags was carried out [14]. The digital photographs with locational information include semantic content which helps searching and indexing. Furthermore, the research proposed many kinds of methods for acquiring location metadata. One of them is using digital cameras or mobile

 ISBN: 978-1-61208-251-6 

phones with built-in GPS chips. So latitude and longitude information is stored in the metadata of image. Another method is to obtain locational information from descriptive text. Most of landscape or scenery image data embedded in web pages generally expresses the locations, buildings or landmarks with text data. Thus, the descriptive text data is used as geographic location tags for the images.

Generally, geocoding is the process of deriving geographic coordinates from geographic data such as addresses or postal codes matching it with reference database. The geocoding goes through three processes which are parsing, matching and locating[10]. Address data is converted to standardized format in the parsing process. And next the input data is matched with pre-built reference database and result records are selected in the matching process. Finally, in the locating process, geographic coordinates of the locations are extracted from the result records. In the processes, various geocoding methods are applied and in terms of the methods, the address matching technique[7], and the 2D and 3D area-based address matching technique[10] have been researched.

In this research, first of all, we concentrated on the indoor environments GPS signal cannot be received accurately. So, instead of using GPS, we chose to use descriptive data having implicit geographic locations. In other words, the descriptive data can be used to derive explicit geographic coordinates in a building. Thus, secondly, during this process, additional process is required for the extraction of descriptive data from images captured by users. So we decided to use OCR engine for extracting words and numbers from store names, room numbers and phone numbers in convention halls and complex malls.

## III. SYSTEM ARCHITECTURE

In this paper, Indoor-space Awareness Architecture is designed as shown in Figure 1.

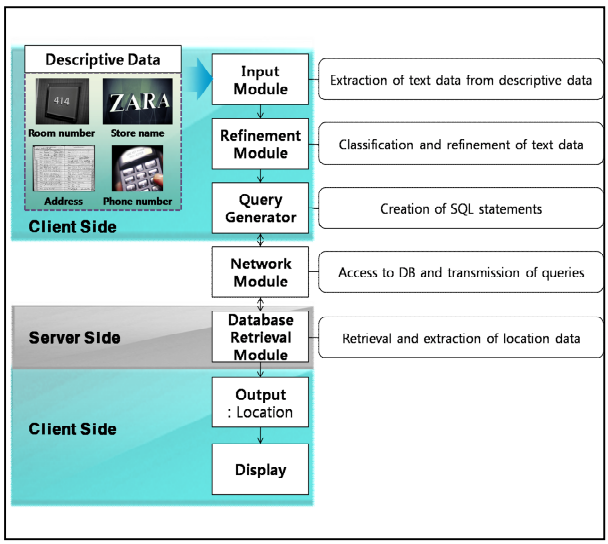



Figure 1. Indoor-space Awareness Architecture.

Among the whole processes, first, the input module extracts text data from descriptive data. Second, the text data is classified into several categories such as an phone number, room number and store name through the refinement module. The text data is parsed to meet the standardized form in order to match it with reference database. And then, query statements are created by the query generator for searching corresponding attributes in database. The query statements are transmitted to the network module, and next retrieval and extraction of location data using the queries are performed in the database retrieval module. The output data is finally displayed in the client side in recognizable form such as polygon or point on a digital map.

### A. *Descriptive Data Input and Refinement*

As a first component shown in left side of Figure 2, the input module receives descriptive data such as store names, phone numbers or room numbers in complex shopping malls or convention halls.

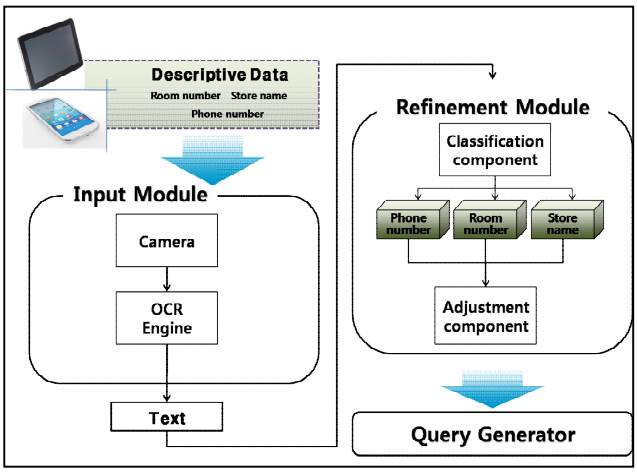



Figure 2. Input and Refinement module of Indoor-space Awareness Architecture.

The camera module captures image data from camera-equipped mobile devices, and then compresses and stores it as image formats. And next, the image data representing descriptive data is sent to the next stage, OCR engine. The OCR engine extracts text data from the image data through several complex processes. OCR is generally divided into on-line and off-line character recognition depending on data acquisition methods. Among them, off-line character recognition is the method that text data is acquired from the fixed static shape of the character in printed documents and signs using a scanner or camera. So it is applied to this research for acquiring text data from images, which represent fixed static shape of characters. Off-line character recognition generally consists of four steps. Firstly, pre-processing adjusts the slant of documents and explicitly divides characters from background for increasing recognition accuracy. Secondly, segmentation is a process to split character strings into many rows, words and characters. Thirdly, a process of feature extraction analyzes the pattern of each character using various methods and then characters

 ISBN: 978-1-61208-251-6 

are individually recognized. Lastly, post-processing increases recognition accuracy taking into account semantic information of context. Incorrectly recognized words are proofread using dictionary data method[5] or HMM method[6]. After the processes of the input module, as a result, the text data is extracted.

As a second stage in the Indoor-space Awareness Architecture, text data is classified into a phone number, room number or store name in the classification component as shown in the right side of Figure 2. The reason for necessity of classification is that it allows the text data to be refined correctly depending on the content types of individual text data. For example, phone numbers consists of a combination of numbers and it is at least greater than minimum digit numbers including area codes depending on many countries without any letters. Afterwards, classified text data is adjusted for the standardized structure. And then, queries are created by the query generator for searching corresponding attributes in database.

### B. *Searching and Display*

Generated queries are transmitted from client side to server side through the network module as shown in Figure 4. The network module connects the client side to the reference database of the server side. It accesses the databases and sends the queries to the database retrieval module. And then, the database retrieval module searches and extracts location data using the queries. In this module, the query manager checks validation of the queries and decides how to access reference database. Next, it is decided that which kind of format the location data extracted from reference DB has in the formatting component. The output data is finally displayed in the client side in recognizable form such as a polygon or point on a 3D digital map.

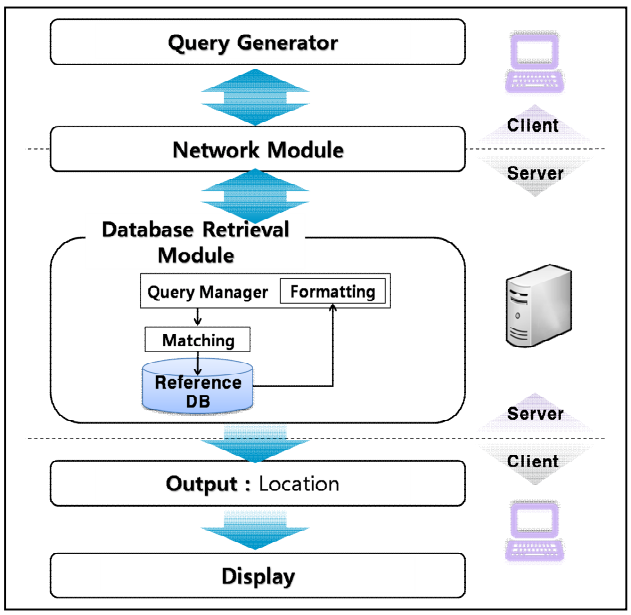



Figure 3. Database Retrieval module and output data of Indoor-space Awareness Architecture.

The reference database schema is designed as shown in Figure 4. The 'Feature' table contains a set of attributes such as an identifier (ID), floor number, room number, coordinates and foreign key(R_ID) which refers to the identifier(ROOM_ID) of 'Room' table. 'Room' table contains a set of attributes such as a phone number, additional number for the same name of stores(ST_ORDER) and foreign key for referring to the identifier of 'RoomName' table. 'RoomName' table is designed including attributes of store names with various languages respectively.

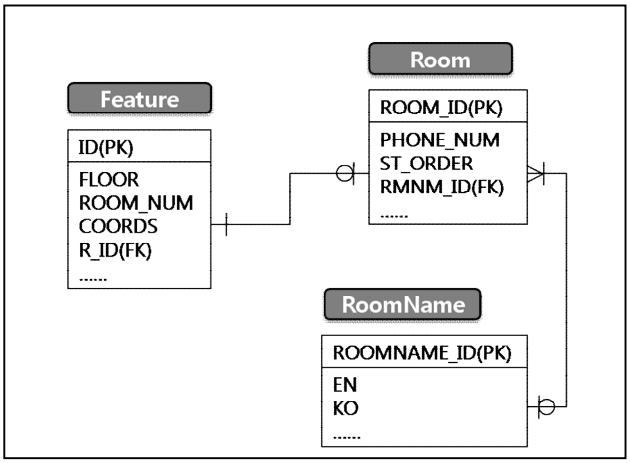



Figure 4. Schema of Reference DB.

Especially, matching process of store names works according to the algorithm below. Firstly, a store name *snt* is directly matched with the list of corresponding attribute. If it is matched at the first attempt, (*x*, *y*, *z*) coordinates are extracted without any additional works. However, if it cannot be matched perfectly, it is partially matched and the location data linked to *output* is extracted. Regardless of the partial matching, if the *output* has candidates or a matched attribute data cannot be found, then matching process operates manually.

```
Input: Boolean b;
       Array output;
Procedure MatchingProcess(Storename snt){
  output←CALL FullMatching(snt)
  IF (output.length = 1 )
    CALL Location(output)
  ELSE IF (output.length = 0 )
    FOR each attribute att from ReferenceDB
      b←CALL IsContain(snt, att)
      IF (b)
        output←att
      END IF
    END FOR
    IF (output.length = 1)
      CALL Location(output)
    ELSE
      CALL ManualMatching(snt)
    END IF-ELSE
  END IF-ELSE
}
```

 ISBN: 978-1-61208-251-6 

## IV. IMPLEMENTATION

Our research is executed based on the Indoor-space Awareness Architecture. A building for the research spot is COEX, which has a largest underground mall and convention centers on the ground. For the implementation of the architecture, an application program is developed based on the android app using Open source OCR program called 'Tesseract'[2][1]. For the actual implementation of the research, as a first stage, sample data is stored in the reference database. As shown in Figure 5 (a), the name of store can be captured as text data using equipped camera in a smart phone. And next, the text data is automatically matched with the store name from the reference database. Then, the current position is represented as a polygon in 3D visualization with information about the store. The keypad enabling manual matching is supported with auto complete text view as shown in (b). For the application of our research, we implemented a query service and the result areas are represented as distinctively colored polygons like (c).

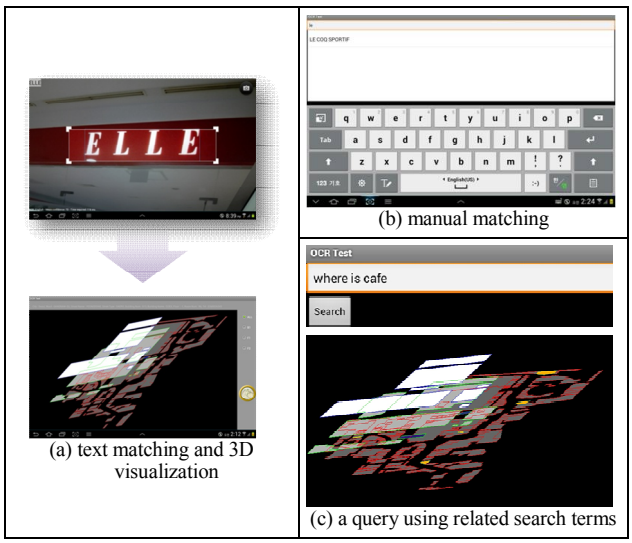



Figure 5. Indoor-space Awareness with 3D Visualization.

Correctly recognizable rate of locations using all kinds of descriptive data is 94.35%. Specifically, the automatic matching rate of English storefront signs is 82.41% using 108 samples, whereas such matching rate of phone numbers is 100% using 10 samples. Although the number of samples is too low in the case of phone numbers, it is analyzed that the error rate in the case of the storefront signs is caused by illumination, various fonts and background color.

## V. CONCLUSION

This research introduced Indoor-space Awareness Architecture using descriptive data for detecting and extracting locational information through modified geocoding method. Among the process, OCR engine is especially applied to recognize text data from the descriptive data. For further researches, special processes should be researched and used in the OCR program for increasing recognition rate. For example, adaptive thresholding needs to be applied for eliminating the effect resulting from illumination. And, for various application of the architecture, the research on incorporating 3D network model can be possible.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2012-0002348).

This work was researched by the supporting project to educate GIS experts.

## REFERENCES

[1] Android-OCR. https://github.com/rmtheis/android-ocr

[2] Tesseract open source OCR engine. http://code.google.com/p/tesseract-ocr.

[3] Baus, J., Kruger, A., Wahlster, W. 2002. A Resource-Adaptive Mobile Navigation System. 2002. International Conference on Intelligent User Interfaces IUI02.

[4] Bekkali, A., Sanson, H., and Matsumoto, M. 2007. RFID indoor positioning based on probabilistic RFID map and kalman filtering. In Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communication, 21-27.

[5] Bokser, M., 1992. Omnifont technologies. Proc. IEEE, vol. 80, pp. 1066-1078.

[6] Chen, M. Y., Kundu A., and Zhou J. 1994. Off-line handwritten word recognition using a hidden Markov model type stochastic network. IEE Trans. Pattern Anal. Machine Intell., vol. 16, pp. 481-496.

[7] Drummond, W.J., 1995. Address matching: GIS technology for mapping human activity patterns. Journal of the American Planning Association, vol. 61, pp. 240-251.

[8] Feldmann, S., Kyamakya, K., Zapater, A., and Lue, Z. 2003. An Indoor Bluetooth-Based Positioning System: Concept, Implementation and Experimental Evaluation. In Proceedings of the International Conference on Wireless Networks, 109-113.

[9] Hays, J., and Efros, A. A. 2008. Im2gps: estimating geographic information from a single image. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition.

[10] Jiyeong Lee, 2009. GIS-based geocoding methods for area-based addresses and 3D addresses in urban areas. Environment and Planning B : Planning and Design, vol. 36, pp. 86-106.

[11] Kaemarungsi, K., and Krishnamurthy, P. 2004. Modeling of Indoor Positioning Systems Based on Location Fingerprinting. In Proceedings of the IEEE INFOCOM, vol. 2, 1012-1022(May 2004).

[12] Li, B., Salter, J., Dempster A. G., and Rizos, C. 2006. Indoor positioning techniques based on wireless LAN. In 1st IEEE Int. Conf. on Wireless Broadband & Ultra Wideband Communications, Sydney, AUS(Mar. 2006).

[13] Mori, S., Suen, C. Y., and Yamamoto, K. Sannella, M. J. 1992. Historical Review of OCR Research and Development. In Proceedings of the IEEE, vol. 80, no. 7, 1029-1058.

[14] Toyama, K., Logan, R., and Roseway, A. 2003. Geographic Location Tags on Digital Images. In Proceedings of the International Conference on Multimedia, 156-166.

 ISBN: 978-1-61208-251-6 

# A Comparison Between Optimization Heuristics of the Best Path Problem Applied to S-Route

Adriano dos Santos, Alex Barradas, Sofiani Labidi, Nilson Costa
Federal University of Maranhão
Sao Luiz, Brazil
{adriano.asr, barradas.alex, soflabidi, nilson2001}@gmail.com

*Abstract* - **The paper contextualizes the traveling salesman problem applied to S-Route, a system developed based upon the ant colony approach and used by the Analysis and Research in Petroleum Analytical Chemistry Laboratory - UFMA in order to generate routes in the process of fuel collection. Some heuristic approaches (constructive and improvement), such as Nearest Neighbor, Clarke and Wright, Mole and Jameson, 2-Opt, 3-Opt and Opt-K, are conceptualized and compared to the Ant Colony. Comparisons between the heuristics, and in some cases the combination of the constructive and improvement occurred through the S-TSP system. Like S-Route, the application was developed in a web environment and integrated with Google Maps API in order to facilitate visualization of the results from georeferenced data. Thus, the essay aims at identifying amongst the listed and/or combined heuristics, the best one regarding cost/benefit to be utilized by S-Route.**

***Keywords***-*Traveling salesman problem; S-Route; Fuel Quality Monitoring; ANP.*

## I. INTRODUCTION

National Petroleum, Natural Gas and Biofuels Agency (ANP in Portuguese) is the Brazilian agency responsible for regulating, supervising and hiring all activities related to petroleum, natural gas and biofuels in Brazil [1]. In order to follow the general indicators of fuel quality traded in Brazil, ANP has the Liquid Fuels Quality Monitoring Program (PMQC in Portuguese), which is summarized in the following stages: Fuel Sample Collection (CAC in Portuguese); Sample Laboratory Analysis, Data Handling and Information Submission to ANP [1, 2].

In this context, the S-route system was developed in order to optimize part of CAC process. Having the principle of obtaining a path from a starting point going through all gas stations selected [2]. However, the systematic arrangement of this scenario can be associated with the graph theory and characterized by the Shortest Path Problem, precisely, the Travelling Salesman Problem(TSP) [3,4].

The purpose of the TSP is to find the lowest total cost of Hamiltonian cycle [5]. However, the TSP is classified as a NP–Complete problem, in other words, the execution time grows exponentially in accordance with the number of points in the route [4, 5]. In this scenario, it is suggested the utilization of heuristic methods that optimize the relationship between time and cost in order to find a solution.

TSP heuristic algorithms are an approach that do not offer the guarantee for the best solution, but seek to meet the standards through a good solution, which approximates the optimal solution and minimize time and cost execution [6].

S-Route system utilizes the ant colony heuristic method to elaborate routes [2]. Aiming to improve the performance of the S-Route system, this essay conducts a comparative study between the ant colony heuristic and other TSP heuristic algorithms: The nearest neighbor; Clarke and Wright (Saving); Mole and Jameson; Ant colony optimization; 2-Opt and 3-Opt e Lin and Kernighan (K-opt).

## II. S-ROUTE SYSTEM

The S-Route System is a prototype of a web-based system that aims to automate part of the process of Fuel Sample Collection, the first phase of PMQC [2]. In State of Maranhão, the Federal University of Maranhão with its Analysis and Research in Petrol Analytical Chemistry Laboratory (LAPQAP / UFMA) is responsible by the PMQC– ANP in monitoring the fuel's quality in State.

The Maranhão State is divided in four regions by the LAPQAP called R1, R2, R3 and R4. This way, the laboratory has one week to collect the samples in each region [2].

The first week of the month is destined to the region R1 (Saint Louis city), so, the initial task is to make 10% (ten percent) of the fuel station in the group of towns of R1. The same thing happens in the others regions, even though needs to be kept the second week to R2, the third one to R3 and the forth one to R4 [2].

Therefore, the first stage of the system is a list of 10% of active gas stations in the data bank and randomly (drawing). Next, the administrator visualizes the layout of the gas stations, using a maps API and then, requests the generation of a route from the listed gas stations.

The first generated route takes into account the savings in time and distance between Federal University of Maranhão (starting point) and the drawn gas stations. Then, the path is displayed on a map, as well as the route description to be trafficked.

S-route application was developed through PHP programming languages [7] and JavaScript [8]. The use of PHP language, for business rules and information management, doesn't present a single or absolute justification in the development, but yet for implementation

 ISBN: 978-1-61208-251-6 

ease. Albeit the use of JavaScript language by the S-Route is not facultative, communication between maps API and application takes place via JavaScript and data structures in XML format.

## III. HEURISTICS

The shortest path problem is originated in the purpose of obtaining the minimum route of an associated path using the graph theory [3, 4, 5]. In this case, a graph may represent a road network and geographical distance from one point to another or from an entire circuit. One of the trends of the shortest path problem is the Traveling Salesman Problem, which represents an optimization problem greatly studied by scholars from several areas, such as: logistics, genetics, production and others [9].

However, the resolution of Travelling Salesman Problem through exact methods or Brute Force algorithms, is not recommended, what is suggested is the utilization of polynomial complexity approximation algorithms that are called by Heuristic methods [10], which allow to obtain reasonable answers to the TSP.

Heuristic methods for TSP, according to the literature, can be classified into two types: Circuit Construction Methods and Circuit Improvement Methods [11].

In the first case, the circuits are built progressively, in other words, the nodes are sequentially inserted in the circuit, under the insertion conditions defined in the algorithm [11]. In the second case, the Circuit Improvement Methods aim to improve the existing Hamiltonian circuit through other methods applied [11].

Hereinafter, the heuristics which were used in the essay development and which comprise both types of methods will be presented.

### A. Nearest Neighbor(NN)

The nearest neighbor heuristic starts with an empty circuit at a starting point in order to seek "the closest" point that is out of the circuit. For each subsequent interaction, the heuristic searches “the closest” point for the last point inserted in the circuit [12].

In summary, the path is constructed as per the shortest distance between these points, in other words, a point is added to a route based upon proximity in relation to the last point inserted. This distance is verified in the matrix, where $d_{ij}$ is the distance between *i* and *j*. The metric applied to the NN approach can refer either on the spatial distance or temporal points.

In short, the NN heuristic is simple to implement, besides achieving good results for short distances, although for long distances it is not so recommended.

### B. Clarke and Wright

Clarke and Wright method (CW) is based upon the concept of "gain" that can be achieved by connecting two knots in succession on a script [13]. The heuristic works similar to the Nearest Neighbor, differentiated by the search for better savings and NN searches the smallest edge.

The savings would be the cost of going and coming back to point 0 going through *a* and *b* without having to go through 0. Instead of going through *a*, getting back to 0, going to *b*, and coming back again to 0.

In essence, the algorithm computes all the savings amongst all pairs of possible vertexes using formula 1:

$$S_{ij} = C_{i0} + C_{0j} - C_{ij} \quad (1)$$

*Sij* represents the path savings of going and coming back to point 0 going through points i and j without having to return to 0 instead of going through i going back to 0, going to j e going back again to 0, as it is suggested by the initial routes that were previously created. *Cij* represents the cost of going from point *i* to *j*.

After calculating all the graph savings, a table of savings is created, with *i, j,* and the savings value. The table lines are ordered up, from the largest to the smallest savings. Then, the path is assembled, by using the vertexes of the table, from the beginning to the end of the table.

The advantage is that Clark and Wright algorithm computational complexity is O(n2), in other words, it is solved in polynomial time. The main contribution of this algorithm can be considered by the fact that it has paved the way to more powerful algorithms that emerged after this one, for example, the Mole and Jameson [14].

### C. Mole and Jameson

Mole and Jameson heuristic [15] is an evolution of Clarke and Wright savings algorithm. The main difference between the two heuristics is the comparisons between the nodes and internal vertexes of the partial path, and allowance of insertions inside this path.

Mole and Jameson algorithm starts from the cost matrix that represents the route relationship, by selecting an initial vertex to build the path. After insertion of the first vertex into the route, the algorithm executes a loop that successively inserts the nodes in the path.

The previous action is performed according to two criteria: proximity and savings. The proximity criterion selects the node that is closer to the current route, according to the two distances calculated by formula 2 [14]:

$$e_{ilj} = C_{il} + C_{lj} - \mu C_{ij} \quad (2)$$

where *C* represents the cost between one vertex to another, *l* is the tested vertex to be inserted, the index *i* represents the beginning of the route *j* means the end of the route.

The savings Criterion selects the best place in the route to insert the chosen *I* vertex. This criterion follows the formula:

$$\sigma_{ilj} = \varphi C_{0l} - e_{ilj} \quad (3)$$

The place that presents the largest savings σ will be selected to receive l. The parameters $\varphi$ and $\mu$ allow changing the behavior of the algorithm in several ways [15].

 ISBN: 978-1-61208-251-6 

The literature indicates to follow the Gaskell criterion [16], where $1 \leq \varphi \leq 2$ e$\mu = \varphi - 1$ , in which the following values $\varphi = 2\ and\ \mu = 1$ are advisable. The algorithm ends when there are no more vertexes to be inserted in the route.

### D. Ant Colony Optimization (ACO)

Ant colony optimization algorithm (ACO) [17] is a constructive meta-heuristic based upon a real behavior of ants using adaptation, cooperation and parallelism techniques [18].

The main idea of this algorithm is the agents indirect communication based on routes trailed by pheromones that are left by ants [17,18] and choice of the best route using the probability.

The ACO is based upon the probability of an ant k being in a point *i* to choose point *j* in an interaction *t*, following the formula below:

$$p_{ij}^{k}(t) = \begin{cases} \frac{[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\Sigma[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}} \\ \kappa \in allowed_{\kappa} \\ 0, others\ cases \end{cases} \quad (4),$$

where the variable α is the pheromone weighting $(0 \leq \alpha \leq 1)$ and β is the heuristic information weighting $(0 \leq \beta \leq 1)$. $\tau_{ij}(t)$ is the pheromone present in the path between *i* and *j*, *t* being updated in every interaction. The value $\eta_{ij}$ represents the ant's attractiveness at point i to visit point j, displayed in the following formula:

$$\eta_{ij} = {}^{1}/_{d_{ij}} \quad (5)$$

The pheromone deposit is calculated in a pheromone matrix $\tau$ , similar to the cost matrix, however the values of this matrix, $\tau_{ij}$, are in constant variation.The utilization of the matrix $\tau$ values occurs after the completion of each route built by ants.

In the update, the pheromone is added to the path as below:

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L_k}, if\ the\kappa - th\ ant\ use\ the\ track(i,j) \\ 0, others\ cases \end{cases} \quad (6)$$

where Q is a project constant and $L_k$ is the length of the circuit of the K-th ant.

The pheromone is also decreased from the edges, simulating the evaporation through the following formula:

$$\tau_{ij}^{k}(t) = (1-\rho)\tau_{ij}^{k}(t) + \Delta\tau_{ij}^{k} \quad (7)$$

where ρ is the pheromone evaporation rate, defined in $(0 < \rho \leq 1)$.

In general terms, k ants are scattered by the nodes starting from the starting node i to j, where the choice of node j will have the highest probability according to formula 4, and thus they carry on building their paths. When all of them complete their paths, the level of pheromone on the paths is updated following the formulas 6 and 7.

Upon completion of the update, a new path search is performed until *t* interaction completion. The process ends with the end of the interactions or when all ants are on the same path.

### E. 2-Opt e 3-Opt Improvement

The 2-Opt and 3-Opt heuristics are the most used and known path improvement methods in the literature. They were proposed respectively by Croes [19] and Lin [20], inaugurating the improvement of k-opt type. These methods work basically by removing k arcs from a script and replacing by other k arcs, with the purpose of reducing the total distance traveled.

To Laporte [21], the higher k value, the better method accuracy, but the higher computational effort. In this context, it highlights the preference of working with two or three arches, therefore, the use of 2-opt and 3-opt heuristics.

The 2-opt method works as follows: from a Hamiltonian path, the first step is to remove two edges from n edges, resulting in a pair of arcs. Then, the arcs are reconnected in an inverted way, the beginning of an arc is connected to the beginning of another arc and the end with the other arc's end, where the process repeats to the other arcs.

In 3-opt case three arcs are considered instead of two in order to assess the alterations in the connections between the nodes, which results in seven possible combinations.

### F. Lin and Kernighan Improvement (LK)

The algorithm proposed by Lin and Kernighan [22] is based upon type k-opt improvements, differentiating itself by the k value variation during the improvement execution. In general terms, the algorithm determines which *k* to be utilized at each interaction without the need to indicate the *k* value before starting the execution.

Basically, the LK algorithm works as follows: The algorithm searches, in increasing values of *k*, which variation results in the shortest route and in each interaction the algorithm performs arc switching.

After a number of r switching, series of tests are performed in order to check if *r +1* must be considered until some stopping condition is satisfied.

Lin and Kernighan [22] defined rules to edge searching and switching, where only sequential switching is permitted, the search for an edge cannot exceed more than 5 nearest neighbors of the current point, and an edge previously excluded cannot be added to the route as well as an edge previously added cannot be excluded.

Starting from the Lin and Kernighan's original idea, other researchers have developed several modifications and improvements, always following the *K* variation line in algorithm execution. One of them was proposed by

 ISBN: 978-1-61208-251-6 

researchers Nguyen, Yoshihara, Yamamori and Yasunaga (LK-NYYY), in the algorithm called LK-NYYY [23].

LK-NYYY algorithm uses a sequence of 5-opt starting movements followed by 3-opt basic movements. By performing this sequence of movements, it's intended to balance the quality of solutions obtained and computational time.

## IV. S-TSP IMPLEMENTATION

In order to find the most suitable heuristics to S-Route, it was necessary to implement each of the ones described previously. The S-TSP was responsible for embodying all heuristics equalizing all input conditions, facilitating the visualization of the results and using the same S-Route technologies [2].

The matching system allows the user to select the amount of gas stations that will be used. Then the next step consists in selecting one of the heuristics, Figure 1 ilustrates the application home screen.


Figure 1. S-TSP.

The first text box represents the distance used to calculate the route, in other words, the sum of sections, in case of present costs in the cost matrix. The second text box is the heuristic execution time expressed in seconds.

Like in S-Route [2], the coordinates are transferred from the database to an XML file and then the JavaScript interacts with the data file. In the XML file there's the identification of each element and related latitude and longitude coordinates.

The class responsible for managing all heuristics, on equal terms, is denominated SuperMap. In this class the gas station sorting operation takes place in a vector in order to visualize the route via Google Maps API.

In order to perform the matching of algorithm performance the following heuristics were implemented in JavaScripts:

- Nearest Neighbor;
- Clark e Wright (Saving);
- Mole e Jameson;
- Ant Colony Optmization
- 2-Opt e 3-Opt;
- Lin e Kernighan(K-opt).

For results validation purposes the Brute force algorithm was also developed.

Each algorithm is represented by a SuperMap class method, where each method has the same parameters, the input criterion is the cost matrix, and the output is the route array. With these methods approach it is possible to use constructive heuristics alongside with improvement heuristics.

Some important descriptions regarding the implementations are listed below:

- Brute Force algorithm is implemented in a recursive performance in order to test all possible combinations of routes and verify which one is the shortest;
- Lin and Kernighan algorithm is implemented in Nguyen, Yoshihara, Yamamori and Yasunaga's technique. However, the algorithm has gone through some modifications in the initial movement sequence from 5-Opt to 3-opt and basic movements from 3-Opt to 2-opt;
- In Ant Colony algorithm implementation, ACO variables used the following values, all of them recommended in the literature [16]: $\alpha$ = 1.0, $\beta$ = 1.0, $\rho$ = 0.1, Number of ants = 10 and = Number of interactions = 10;
- All the other heuristics followed chapter III descriptions.

## V. COMPARATIVE AND RESULTS

In order to find the best heuristic comparisons were carried out in two stages. At first the best constructive and improvement heuristics are verified independently. In the second stage, the best constructive algorithms as well as improvement algorithms are combined and compared.

### *A. First Matching Stage.*

The first testing stage aims to find the best improvement and construction heuristic. At this stage, the heuristics are compared to each other by utilizing the total distance as per table 1.

 ISBN: 978-1-61208-251-6 

TABLE I. CONSTRUCTIVE HEURISTICS

| | 10 stations | 15 stations | 20 stations | 25 stations | 30 stations |
|---|---|---|---|---|---|
| **TSP** | 33.463 | 45.609 | 50.574 | 72.107 | 62.989 |
| **Clark and Wright** | 36.235 | 45.589 | 54.532 | 63.379 | 66.306 |
| **M&J** | 34.244 | 43.576 | 43.878 | 51.288 | 54.514 |
| **ACO** | 32.637 | 42.384 | 43.726 | 52.536 | 55.394 |
| **BruteForce** | 32.637 | 42.384 | - | - | - |

Comparing Table 1, the best results based upon distance are found by the Ant Colony (ACO) and with similar performance to it, the Mole & Jameson (M&J). The execution time was disregarded due to the fact that all the heuristics performed below 1 second.

In Table 2 the improvement heuristic performance is verified. For all improvement algorithms, the test uses as input the same array of the ordered gas stations.

TABLE II. IMPROVEMENT HEURISTICS

| | 10 stations | 15 stations | 20 stations | 25 stations | 30 stations |
|---|---|---|---|---|---|
| **2-OPT** | 33.051 | 49.839 | 52.867 | 65.011 | 71.742 |
| **3-OPT** | 33.051 | 48.319 | 51.347 | 64.598 | 68.036 |
| **Lin and Kernighan** | 32.637 | 45.39 | 50.416 | 58.912 | 61.698 |
| **BruteForce** | 32.637 | 42.384 | - | - | - |

A performance better than Lin-Kernighan in 2-OPT and 3-OPT is noticed. Out of the OPT heuristics, the one that had the best performance was 3-OPT, which is justified by the greater amount of switching cycles.

However, Lin and Kernighan has a higher cost when compared to 2-OPT and 3-OPT, as you can see in Table 3.

TABLE III. EXECUTION TIME OF IMPROVEMENT HEURISTICS

| Time/sec | 2-OPT | 3-OPT | Lin and Kernighan |
|---|---|---|---|
| **10 stations** | 0.00199 | 0.00399 | 0.02399 |
| **15 stations** | 0.00299 | 0.01900 | 0.06599 |
| **20 stations** | 0.00600 | 0.02599 | 0.32200 |
| **25 stations** | 0.02000 | 0.07290 | 1.45299 |
| **30 stations** | 0.01100 | 0.10899 | 412.800 |

For the second matching stage the chosen improvement heuristics, picked from Table 2 and Table 3 results, were Lin and Kernighan and 3-OPT.

### *B. Second Matching Stage.*

The second testing stage aims to find the best heuristic by verifying simple heuristics as well as combined heuristics. The combined heuristics are the construction ones which have their results improved by the improvement heuristics.

The heuristic defined in the first stage were tested with a larger amount of gas stations based upon distance as shown in table 4.

TABLE IV. HEURISTIC MATCHING

| Stations | M&J | M&J + 3OPT | M&J + LK | ACO | ACO + 3OPT | ACO + LK |
|---|---|---|---|---|---|---|
| **20** | 43.878 | 42.930 | 42.686 | 43.726 | 42.930 | 42.686 |
| **30** | 54.514 | 53.565 | 52.725 | 55.394 | 54.927 | 53.640 |
| **40** | 74.352 | 71.417 | 64.667 | 68.300 | 67.595 | 67.595 |
| **50** | 81.876 | 77.614 | 68.693 | 70.560 | 70.291 | 70.139 |
| **60** | 104.121 | 100.168 | 88.927 | 84.339 | 83.666 | 83.490 |
| **80** | 155.866 | 144.459 | 123.951 | 111.439 | 110.933 | 110.489 |
| **100** | 175.055 | 163.648 | 134.412 | 122.840 | 121.074 | 116.751 |

M&J obtained better performance over ACO regarding the preparation of route for 25 gas stations as shown in Table 1 and 30 gas stations as shown in table 4. ACO succeeded in constructing the route to the quantity of 10, 15, 20, 40, 50, 60, 80 and 100 gas stations (see Table 1 and Table 4).

With the combination of 3-OPT improvement heuristics and constructive heuristics (M&J and ACO), ACO continues to obtain better results over M&J except when the amount is equal to 30 gas stations.

By combining L&K to M&J and ACO heuristics, respectively, M&J showed better results (30, 40 and 50) than ACO except when the amount of gas stations was equal to 60 gas stations.

Table 5 shows the execution time of the heuristics as follows:

TABLE V. EXECUTION TIME(SEC) OF HEURISTICS

| Stations | M&J | M&J + 3OPT | M&J + LK | ACO | ACO + 3OPT | ACO + LK |
|---|---|---|---|---|---|---|
| **20** | 0.004 | 0.031 | 0.3329 | 0.025 | 0.0569 | 0.3209 |
| **30** | 0.023 | 0.027 | 41.789 | 0.032 | 0.069 | 4.210 |
| **40** | 0.037 | 0.207 | 15.700 | 0.085 | 0.230 | 23.200 |
| **50** | 0.070 | 0.500 | 101.500 | 0.052 | 0.480 | 102.290 |
| **60** | 0.088 | 0.976 | 300.430 | 0.115 | 1.006 | 301.530 |
| **80** | 0.223 | 2.790 | 1623 | 0.159 | 2.782 | 1731 |
| **100** | 0.364 | 7.935 | 7603 | 0.277 | 7.506 | 7570 |

In general, M&J has a lower computational cost ACO both in its original state as combined with the improvement heuristics. Albeit, the results do not show significant differences when they are compared to each other.

In this context, the combinations of L&K heuristics and constructive heuristics presented the best results in distance criterion albeit the execution time is high when compared to 3-OPT.

## VI. CONCLUSION

The research addressed the traveling salesman problem applied to the S-Route system [2]. The objective was to

 ISBN: 978-1-61208-251-6 

compare the heuristics used originally by the S-Route, to ACO to other relevant heuristics in the literature.

Thus, the essay selected the following constructive heuristics: Nearest Neighbor, Clark and Wright (Saving); Mole and Jameson. For heuristic improvement, the research addressed: 2-OPT and 3-OPT; Lin and Kernighan (K-OPT).

The combination amongst L&K M&J and ACO heuristics, respectively, achieved the best results related to the distance.

However, the computational cost of L&K combinations was very high when compared to 3-OPT combinations, which somehow do not favor the utilization of L&K. In turn, the combination between 3-OPT + ACO got closer to the results of L&K + M&J and L&K + ACO, which did not occur to 3-OPT + M&J.

In this context, the essay suggests, for S-Route implementation purpose, the combination between ACO 3-OPT heuristic considering the cost benefit between distance and execution time.

## ACKNOWLEDGMENT

Thank you very much to everyone who helped directly and indirectly and allowed the development of this research. Thanks in particular to PRH39-ANP and PPGEE-UFMA programs, and to LSI, LAPQAP and LPQA laboratories.

## REFERENCES

[1] ANP, "Programa de Monitoramento", Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, 2011. Accessed september 26, 2011. Available at http://www.anp.gov.br/?pg=33970

[2] A. Barradas, A .Santos, S. Labidi, and N. Costa, "A Heuristic Approach Based on the Ant Colony Optimization for the Routes Elaboration on the Fuel Collection for the Brazilian Petrolium Agency" IARIA – GEOProcessing 2011, pp. 69-74, ISBN: 978-1-61208-118-2.

[3] N. Aras, B. J. Oommen, and I. K. Altinel, "The Kohonen network incorporating explicit statistics and its application to the traveling salesman problem". Neural Networks, 12rd ed, vol. 9, 1999, pp. 1273–1284.

[4] K. Helsgaun, "An effective implementation of the Lin-Kernigham Traveling Salesman Heuristic", European Journal of Operational Research, vol. 126, 2000, pp.106-130.

[5] C. H. Papadimitriou, "The Euclidean travelling salesman problem is NP-complete". Theoretical Computer Science, 4, 1978, pp.237–244.

[6] P. H. Siqueira, M. T. A. Steiner, S. Scheer, "A new approach to solve the traveling salesman problem". Amsterdam: ScienceDirect - Neurocomputing: A new approach to solve the traveling salesman problem, 2006.

[7] PHP, Manual, 2013. Accessed september 23, 2013. Available at http://www.php.net/manual/pt_BR/preface.php.

[8] J. Resig, "Pro JavaScript Techniques", Apress, 2006.

[9] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook "The travelling salesman problem: a computational Study". Princeton: Princeton University Press, 2006. ISBN 978-0-691-12993-8.

[10] A. Diaz, M. Laguna, P. Moscato, F.T. Tseng, F. Glover, and H. M. Ghaziri, "Optimización Heurística y Redes Neuronales". Editorial Paraninfo, Madrid. 1996.

[11] D. E. Rosenkrantz, R. E. Stearns & P. M. Lewis II, "An analysis of several heuristics for the traveling salesman problem". SIAM J. Comput., vol. 6, 1977, pp. 563-581.

[12] M. M. Solomon, "Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints". Operations Research, vol. 35, ed. 2, 1987, pp. 254-265.

[13] G. Clarke, and J. Wright, "Scheduling of vehicles from a central depot to a number of delivery points", Operations Research, vol. 12 ed. 4, 1964, pp. 568-581.

[14] M. R. Heinen, and F. S. Osório, "Algoritmos Genéticos Aplicados ao Problema de Roteamento de Veículos". Hífen, Uruguaina. vol. 30, ed.58, 2006.

[15] R. H. Mole, and R. S. Jameson, "A sequential routing-building algorithm employing a generalized savings criterion", Opl. Res Q, vol. 27, 1976, pp. 503–512.

[16] T. J. Gaskell, "Bases for the vehicle fleet scheduling", Opl. Res. Q, ed. 18, 1967, pp. 281–294.

[17] M. Dorigo, "Optimization, Learning and Natural Algorithms". M.. PhD thesis, Italy: Politecnico di Milano, 1992.

[18] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents". IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics vol. 26 ed.1, 1996, pp. 29–41.

[19] G. A. Croes. "A method for solving traveling salesman problems". In: Operations Research n. 6, 1958, pp. 791-812.

[20] Lin, S. "Computer solutions of the traveling salesman problem". In: Bell System. Technical. Jour nal. n. 44, 1965, pp. 2245-2269.

[21] G. Laporte, H. Mercure, and Y. Nobert, "A Branch and Bound algorithm for a class of asymmetrical vehicle routing problems". Journal of the Operational Research Society. vol. 43, ed. 5, 1992, pp. 469-481.

[22] S. Lin, and B. Kernighan, "An Effective Heuristic Algorithm for the Traveling Salesman Problem". Operational Research, vol. 21, 1973, pp. 498-516.

[23] H. D. Nguyen, I. Yoshihara, M. Yamamori, "Implementation of an Effective Hybrid GA for Large-Scale Traveling Salesman Problems". IEEE Transaction on System, Man, and Cybernetics-PART B: Cybernetics, vol. 37, no. 1, 2007, pp. 92-99.

 ISBN: 978-1-61208-251-6 