GEOProcessing 2011

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

February 23-28, 2011 - Gosier

Guadeloupe, France

GEOProcessing 2011 Editors

Claus-Peter Rückemann, WWU Münster / Leibniz Universität Hannover / HLRN, Germany

Ouri Wolfson, University of Illinois - Chicago, USA

GEOProcessing 2011

Foreword

The Third International Conference on Advanced Geographic Information Systems, Applications, and Services [GEOProcessing 2011], held between February 23-28, 2011 in Gosier, Guadeloupe, France, brought 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. As an example, they can be used for assessment of accidents from chemical pollution by considering hazardous chemical zones dimensions represented on a computer map of the region's territory.

Geographical sensors and satellites provide a huge volume of spatial data which is available on the Web. Making use of Web Services, the users are able for provisioning and using these services instead of only for document searching. These services are published in a directory and may be automatically discovered in a given context by software agents. Accessing large digital geographical libraries with geo-spatial information raises some challenges with respect to data semantics, interfaces, data accuracy and updates, distributed processing, as well as with discovery, indexing and integration of geographical information systems; this raise the issue of distributed catalogs forming a federation of spatial databases. Some spatial data infrastructures use service-oriented architecture for accessing these large databases via Web Services.

We take here the opportunity to warmly thank all the members of the GEOProcessing 2011 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 GEOProcessing 2011. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the GEOProcessing 2011 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that GEOProcessing 2011 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the areas of geographic information systems, applications and services.

We are convinced that the participants found the event useful and communications very open. We also hope the attendees enjoyed the beautiful surroundings of Gosier, Guadeloupe, France.

GEOProcessing 2011 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
Rifaat Abdalla, Defence R&D Canada - Toronto, Canada
Ouri Wolfson, University of Illinois - Chicago, USA

GEOProcessing 2011

Committee

GEOProcessing Advisory Committee

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
Rifaat Abdalla, Defence R&D Canada - Toronto, Canada
Ouri Wolfson, University of Illinois - Chicago, USA

GEOProcessing 2011 Technical Program Committee

Rifaat Abdalla, Defence R&D Canada - Toronto, Canada
Riccardo Albertoni, Consiglio Nazionale delle Ricerche - Genova, Italy
Thierry Badard, Université Laval - Québec, Canada
Fabian Barbato, Universidad de la Republcia - Montevideo, Uruguay
Brandon Bennett, University of Leeds, United Kingdom
Michela Bertolotto, University College Dublin, Ireland
Ling Bian, University at Buffalo, USA
Lorenzo Bigagli, Institute of Methodologies for Environmental Analysis of the National Research Council (IMAA-CNR), Italy
Morten Borrebæk, Norwegian Mapping Agency, Norway
Carsten Brockmann, Universität Potsdam, Germany
Boyan Brodaric, Geological Survey of Canada, Canada
Christophe Claramunt , Naval Academy Research Institute, France
Eliseo Clementini, University of L'Aquila, Italy
Alfredo Cuzzocrea, Italian National Research Council & University of Calabria - Rende (CS), Italy
Chenyun Dai, Purdue University, USA
Maria Luisa Damiani, Università degli Studi di Milano, Italy
Cláudio de Souza Baptista, University of Campina Grande, Brazil
Monica De Martino, Consiglio Nazionale delle Ricerche - Genova, Italy
Anselmo C. de Paiva, Universidade Federal do Maranhão, Brazil
Antonios Deligiannakis, Technical University of Crete, Greece
Jean-Yves Delort, Macquarie University - Sydney, Australia
Petre Dini, Concordia University, Canada / IARIA, USA
Javier Dominguez, CIEMAT - Science & Innovation Ministry, Spain
Suzana Dragicevic, Simon Fraser University- Burnaby, Canada
Betsy George, Oracle Corp., USA
Diego Gonzalez Aguilera, University of Salamanca - Avila, Spain
Björn Gottfried, University of Bremen, Germany
Enguerran Grandchamp, Université des Antilles et de la Guyane, Guadeloupe
Carlos Granell Canut, Universitat Jaume I- Castellón, Spain
Cory Henson, Wright State University / Kno.e.sis Center, USA
Gerard B.M. Heuvelink, Wageningen University and Research Centre, The Netherlands
Jerry Hobbs, University of Southern California / ISI, USA

Erik Hoel, Environmental Systems Research Institute - Redlands, USA
Zhou Huang, Peking University - Beijing, China
Prateek Jain, Wright State University - Dayton, USA
Vana Kalogeraki, Athens University of Economics and Business, Greece
Baris M. Kazar, Oracle Corp., USA
Jaroslav Klokocník, Academy of Sciences of the Czech Republic - Fricova, Czech Republic
Joel Langlois, BRGM, France
Fabio Luiz Leite Júnior, Universidade Estadual da Paraíba, Brazil
Weidong Li, Huazhong Agricultural University - Wuhan, China
Ling Liu, Georgia Institute of Technology, USA
Qing Liu, CSIRO, Australia
Victor Lobo, New University of Lisbon/Portuguese Naval Academy, Portugal
Qifeng Lu, MacroSys LLC. - Arlington, USA
Hervé Martin, University of Grenoble - St.-Martin d'Hères, France
Nicola Masini, Institute for Archaeological and Monumental Heritage / University of Basilicata - Potenza, Italy
Patrick Maué, University of Muenster, Germany
Dunja Mladenic, JSI, Slovenia
Adrian Mocan, SAP AG / SAP Research CEC Dresden, Germany
Stephan Mäs, Technische Universität Dresden, Germany
Silvia Nittel, University of Maine - Orono, USA
Daniel Orellana Vintimilla, Wageningen University, The Netherlands
Olaf Østensen, Norwegian Mapping Agency, Norway
Michael Pantazoglou, NKUA, Greece
Edzer Pebesma, University of Münster, Germany
Ross Purves, University of Zurich - Irchel, Switzerland
Bernd Resch, Massachusetts Institute of Technology/Senseable City Lab - Cambridge, USA
Kai-Florian Richter, The University of Melbourne, Australia
François Robida, BRGM, France
Dumitru Roman, SINTEF, Norway
Claus-Peter Rückemann, WWU Münster / Leibniz Universität Hannover / HLRN, Germany
Markus Schneider, University of Florida, USA
George Spanoudakis, City University London, UK
Kathleen Stewart, The University of Iowa, USA
Eldar Sultanow, Universität Potsdam, Germany
Juergen Symanzik, Utah State University - Logan, USA
Naohisa Takahashi, Nagoya Institute of Technology, Japan
Vlad Tanasescu University of Edinburgh, UK
Ergin Tari, Istanbul Technical University, Turkey
Jean-Claude Thill, University of North Carolina at Charlotte, USA
Ioan Toma, STI Innsbruck, Austria
Theodore Tsiligiridis, Agricultural University of Athens, Greece
Pablo Fernando Vanegas Peralta, Katholieke Universiteit Leuven, Belgium
Hans Voss, Fraunhofer Institute Intelligent Analysis and Information Systems (IAIS) - Berlin, Germany
Jue Wang, Washington University - Saint Louis, USA
Nancy Wiegand, University of Wisconsin-Madison, USA
Eric B. Wolf, US Geological Survey - Boulder, USA
Ouri Wolfson, University of Illinois - Chicago, USA

Ningchuan Xiao, The Ohio State University - Columbus, USA
Chuanrong Zhang, University of Connecticut - Storrs, USA
Wenbing Zhao, Cleveland State University, USA

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

Integrating Spatial Information into JSF Java EE Web Applications with *GeoJSF*

Thorsten Kisner, Helge Hemmer and Klaus Jacobi
AHT GROUP AG
Management & Engineering
Essen, Germany
Email: {t.kisner,h.hemmer,jacobi}@aht-group.com

***Abstract*—In recent years an increasing acceptance of geographical information systems (GIS) to be used by clients in their web browsers and therefore a high demand for such WebGIS systems is observable. The processing and visualization of spatially referenced data became more and more important, even in fields of applications that are not traditional GIS domains like earth observation or remote sensing. Management information systems (MIS) or social networks among others have also evolved requirements for GIS features. For implementing enterprise architectures, many powerful frameworks and (Open Source) products are available on the market; considering the Java programming language in its enterprise edition (JavaEE) for example, several compliant application servers can be used, along different solutions to share geospatial data and many ways to work with these data in the clients web browser. Being in this concrete scope of developing a Java EE software with a presentation layer implemented in Java Server Faces (JSF), which is part of the Java EE standard specification, one quickly realizes the lack of a comprehensive integration of GIS functionality into JSF. In this paper, we propose a framework to develop web applications with geospatial elements by an own implementation of a component library for JSF technology, enabling the developer to enrich his application with powerful GIS features at ease.**

***Keywords*-WebGIS; WMS; WFS; JSF; JavaEE; integration; component library; spatial data**

I. Introduction

Compared to traditional desktop geographical information systems (GIS), a web based GIS based Open Source tools has several advantages. WebGIS solutions have a much broader accessing scope, the system can be accessed without the need of software installations, maintenance and licensing on the client side. Client software is (mostly) independent from platform and architecture, only a modern[1] web browser is required to access the information. Commercial GIS products with licensing costs and annual maintenance fees like *ArcIMS* offer state of the art technology with a rich variety of features, even if only a limited number of functionalities is used. If existing applications should be extended only with standard GIS functionalities or just a simple map, commercial GIS suites might be an oversized solution.

Either way – out-of-the-box solutions always require individual customizing. In both cases, using commercial or open source based software with free available libraries and products, detailed programming skills are required.

In the developing domain of JavaEE applications with JSF for it's presentation layer, developers are used to include existing (component) libraries into their applications to add the required features. For WebGIS applications, a JSF component library offering a seamless integration with features of user interaction would be highly appreciated. Such a framework is presented in this paper.

The remainder of this paper discusses the related work and positions the proposed approach (Section II). Section III introduces the Java Server Faces framework in which domain our proposal is implemented followed by a discussion of required GIS standards in Section IV. The implementation itself is discussed from a technical point of view in Section V. Section VI presents results and ends with an overview on future work.

II. Related Work

In many software domains, classic desktop application are replaced by rich internet applications (RIA). These products are running on a server and are accessed by clients through their Internet browser, giving them nearly the same user experience than in a desktop system. The same applies to geographical information systems (GIS). When common users of todays information systems come across the keyword "geographical information", most associate this with *Google Maps*, a web based GIS launched by Google in 2006. It is an obvious thought, because simple GIS features in web sites are often realized through *Google Maps*.

Even more complex information systems are build on top *Google Maps* as a viewer for geo-spatial data. In [1], it is used, besides other technologies, to show water resources on a map. Data is processed inside the application and visualized using a KML (Keyhole Markup Language) file through the *Google Maps API*. In other scenarios, information systems need to use their own maps and advanced features which *Google Maps* does not offer. Especially when it comes to infrastructure provision, there are often higher requirements to the GIS functionality.

Therefore, a lot of work was done to implement custom

[1]With activated JavaScript or ActiveX.

 ISBN: 978-1-61208-118-2

WMS/WFS[2] clients. For example, a system to monitor and plan pipelines of a urban drainage network is described in [2]. The GIS component of their project is realized through a Java Applet that is included on a web page.

Meeting the requirements of nowadays software projects, recent developments are more and more integrating AJAX (Asynchronous JavaScript and XML) features to the GIS functionality. Whereas [3] is combining *Google Maps* with extra information from Geo Web Services, a "Geo Stack" of several tools with a front-end utilizing the *OpenLayers* library was built up in [4]. The capabilities of *OpenLayers* are inspected in [5], pointing out that it can replace a classical desktop GIS client.

Geographical data processing is likely being only one component of a comprehensive information system, therefore it is needed to be integrated in larger infrastructures. One approach is a Service Oriented Architecture (SOA), like it is done in [6]. The implementation presented there is accessing WMS/WFS servers through XML Web Services. Respectively in [7], where JavaEE patterns are used to develop a system with an Enterprise Service Bus. A SOA environment in conjunction with a service to offer Styled Layer Definition files for *OpenLayers* is presented in [8]. A SOA style service bus is presented in [9] for sensor data in an application of Tsunami warning.

Besides these approaches to integrate applications as parts in a wide infrastructure, [10] concentrates on the joining of spatial data of different sources in one map using open source software like *OpenLayers* and *Mapserver*.

Another kind of integration is proposed by [11] using the commercial JSF Application Development Framework (ADF) by ESRI to create WebMap applications inside web front-ends. The same client is also used in many other web GIS like NASAs *Apollo Analysts Notebook* [12]. Besides *OpenLayers* and the ESRI product, there are other options like *MapXTreme*, which is inspected in [13].

The approach presented in this paper is extending the ideas of *geo-faces*[3] and *ol4jsf*[4], both projects providing a JSF component to integrate the *OpenLayers* client into a JavaEE web application. Like *geo-faces*, our component library *GeoJSF* uses the RichFaces Component Development Kit (CDK) – a toolkit to simplify the implementation of custom JSF components.

Whereas *geo-faces* and *ol4jsf* are basically offering a JSF tag to initialize an OpenLayers instance, *GeoJSF* is extending this approach with an integration framework and offers AJAX based interactions between the JavaScript client running on the client side and the JavaEE system on the server side with the additional integration of Enterprise Java Beans (EJB).

[2]WMS and WFS is discussed in Section IV.

[3]http://code.google.com/p/geo-faces

[4]https://ol4jsf.dev.java.net

Other proposals like [14], which is trying to find a universal information representation for geographical data, are a good option when having a heterogenous application environment. A novel approach to query spatial data is presented in [15], but concluding that object-relational mapping is the method of choice in typical JavaEE systems.

For other technologies not residing in the JavaEE framework, [16] presents novel ideas for implementing a web based application with techniques like Scalable Vector Graphics (SVG), whereas [17] presents an overview to build up a geo portal using Open Source technology.

III. Java Server Faces

Java Server Faces (JSF) is a server-side framework to develop the web enabled presentation layer of an enterprise application. Developers can use predefined and reusable component libraries (like the one we describe in our paper), link component elements to data sources managed by the server and connect client-side events to a server-side event handler.

The reference implementation for JSF is included in the Java Specification Request 127 [18] and includes simple GUI elements like links, buttons or input fields which have direct equivalents to tags of the HTML specification.

The components included in the reference implementation are sufficient to create web applications, but the web interface will only appear in an old fashioned style like web sites in the 90's. Modern web sites or rich internet applications require high interactivity and have to satisfy graphical layouts ("Web 2.0"). There are many sophisticated components libraries available like *ICEfaces* from ICEsoft, *RichFaces* from Red Hat or *Trinidad* from the Apache Software Foundation which fulfill these s requirements.

A. JSF Architecture

The component framework is a Model-View-Controller (MVC) architecture shown in Figure 1.

The *Controller* is responsible for the navigation and user interaction and is implemented in the `FacesServlet`. Developers can define navigation rules and assign event handlers to different components.

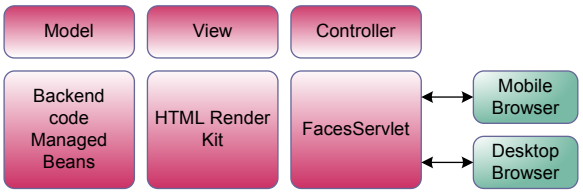


Figure 1. MVC architecture with JSF

The *Model* is represented by "Managed Beans" and is the Java part of JSF. Managed Beans are invoked by *View* components and are implemented in a POJO[5] concept. The

[5]Plain Old Java Object

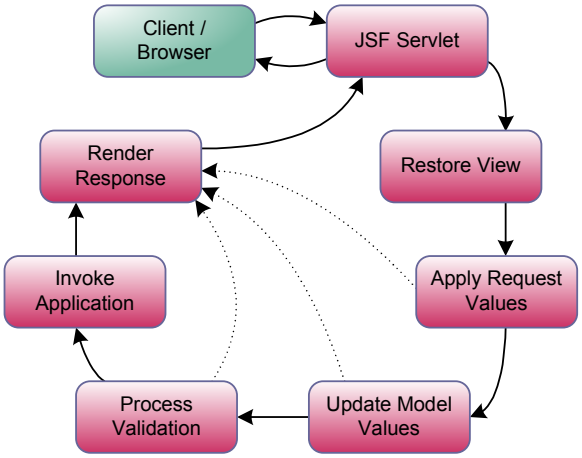


Figure 2. Life Cycle of a JSF application (short-circuited by errors)

View components itself are provided the by components libraries.

B. JSF Life Cycle

The different phases of a JSF request/response turn are outlined in Figure 2.

The state of the component tree is restored from the previous request in the *Restore View* phase, in the next phase *Apply Request Values* all new values are applied to the corresponding components. Before the model values are updated, validations and converters are invoked in phase three. After updating model values, the phase *Invoke Application* will be entered and after processing of all events, the response is rendered and sent back to the client.

There are shortcuts available from the first phases to *Render Response*. They are called if no query data are available or errors in the validation or conversion phase occur.

C. Implementation of components

A JSF component is a set of Java classes and XML configuration files. To develop a component, the followings steps are required and undertaken in our component library:

1) Implementation of a Java class which extends the basic component classes of JSF.
2) Implementation of a renderer class for the default render kit of the reference implementation which creates the output of the specific class.
3) Implementation of a class which describe the tags for the JSP[6] class.
4) Definition of the *Tag Library Definition* (TLD) file. This file describes all available tags with their required or optional attributes and allowed types.
5) The renderer of each component is defined in the *JSF Configuration* file. For each component the family (defined in the JSF standard), a Uniform Resource Identifier (URI) for the component and the actual renderer class defined above must be specified.

All elements, the three individual Java classes per element and both configuration files, will be provided by our JSF library as a JAR archive which directly can be used by web application developers. The archive must be available in the class path of the servlet container and the library must be declared in the head of the JSF file like shown in Listing 1.

Listing 1. Usage of *GeoJSF*

```
<?xml version="1.0" encoding="utf-8" ?>
<jsp:root version="2.1"
  xmlns:html="http://www.w3.org/1999/xhtml"
  xmlns:jsp="http://java.sun.com/JSP/Page"
  xmlns:f="http://java.sun.com/jsf/core"
  xmlns:h="http://java.sun.com/jsf/html"
  xmlns:a4j="http://richfaces.org/a4j"
  xmlns:rich="http://richfaces.org/rich"
  xmlns:geojsf="http://geojsf">
<jsp:output doctype-root-element="html"
  doctype-public="-//W3C//DTD XHTML 1.1//EN"
  doctype-system=
      "http://www.w3c.org/TR/xhtml11/DTD/xhtml11.dtd" />
<jsp:directive.page contentType="text/html; charset=utf-8"
                    language="java" />
<html>
  <jsp:directive.include file="../include/head.jspx"/>
  <body><f:view><div><rich:layout>
  <rich:layoutPanel position="top"><h:form><rich:panel>
<geojsf:map id="olMap" serviceName="LCBC"
  overviewMapDiv="dOverview"
  width="500" height="400"
  serviceURL="#{MapsBean.ol.wmsLayer.value.url}"
  layerOptions="#{MapsBean.ol.wmsLayer.value.params}"
  centerX="#{MapsBean.ol.view.x}"
  centerY="#{MapsBean.ol.view.y}"
  zoomLevel="#{MapsBean.ol.view.zoom}">
</geojsf:map>
  </rich:panel></h:form></rich:layoutPanel>
  </rich:layout></div></f:view></body>
</html>
</jsp:root>
```

The map is included with the `<geojsf:map/>` tag and most of the attributes are bound to the managed bean `MapsBean` with the JSF Expression Language.

D. JSF Expression Language

JSF-EL is similar (but not compatible) with JSP-EL and is processed in the *Update Model Values* and *Render Response* phases (see Figure 2). The navigation to object properties is analogue to the XML Path Language (XPath) and is embedded in the syntax `#{...}`. The expression itself can be object-value bindings, arithmetic or logical expressions and method bindings.

IV. Geographical Information Systems

A geographic information system (GIS) (aka geospatial information system) is a system that captures, stores, analyzes, manages, and presents data that are linked to location.

The most important standards in this area are the ISO series 191xx[7] and publications from the Open Geospatial Consortium (OGC) [19].

[6]Java Server Pages

[7]ISO 19107, 19109, 19111, 19115, 19136

 ISBN: 978-1-61208-118-2

The OGC has developed different specifications for data exchange, including the Web Map Service (WMS) to encode maps as images, the Web Feature Service (WFS) for working with data referenced by geographic objects or vector data, the Web Coverage Service (WCS) for continuous data, live access to observations from sensors with the Sensor Collection Service (SCS) and the Geographic Markup Language (GML) to encode geographic objects and linked data for transport in XML data.

Due to the relevance of WMS and WFS to *GeoJSF* these standards are discussed in detail in the following sections.

A. Web Map Service

Being a special case of a Webservice, a Web Map Service (WMS) is an interface specification published by the OGC [20] to request and serve map images in the internet. An OGC compliant WMS uses Hypertext Transfer Protocol (HTTP) for transport and has to implement three functionalities.

A client can ask with `GetCapabilities` for general properties of the server like available output formats[8] or available layers for a specific map. The response is send to the client in a XML document.

The concrete map images of a spatial referenced map are requested with a `GetMap` query. The query can include optional parameters like the geospatial reference system, image size, map section or output format. A map image can directly requested by any web browser (wiht plain HTML), but intermediate clients (e.g. JavaScript) offer possibilities to change layers, zoom or move the map.

The optional `GetFeatureInfo` answers queries to specific coordinates or selected areas (with an optional search radius) in the actual map section. The response contains thematic information of the map (layer). The response can be offered in XML or formatted in HTML.

B. Web Feature Service

The Web Feature Service (WFS) [21] defined by OGC specifies an interface to access spatial data on distributed GIS systems. Other than the WMS, only access to vector data is possible.

OGC compliant WFS server support six functionalities, use HTTP for transport and encode the response with the XML based Geography Markup Language (GML).

The `GetCapabilities` response contains information on available feature types and supported operations. `DescribeFeatureType` provides detailed information on the structure of a feature type.

The spatial referenced data itself are requested by `GetFeature`, the response object can be filtered with given feature types or spatial dependencies. The GML encoded response object can be used to create a XLink[9] query to access further elements with `GetGmlObject`.

Write access to the provided data is provided with `Transaction`, this supports a revertable transaction to create, modify and delete objects. The `LockFeature` operation during a transaction prevents a concurrent write access to the specified object by other instances.

V. The JSF component library GeoJSF

A. Architecture

In our scenario we are dealing with a distributed service architecture and different domains of information processing like shown in figure 3.

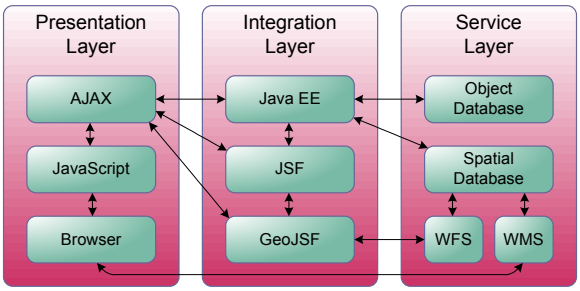


Figure 3. Layered architecture and components

- *Browser (Java Script):* The client side logic is located in JavaScript code running in a web browser. JavaScript is responsible for the client side event handling and the execution of server side commands. Data is transferred between client and server by JavaScript functions utilizing the AJAX framework. This allows partial page rendering and a high response time of the application.
- *Java EE Application Server:* The Enterprise Java Bean (EJB) specification is extended by spatial data types using HibernateSpatial. With this extension objects can be directly modelled on the EJB side with spatial characteristics like points, polygons or multi-polygons.

Listing 2. EJB annotation for spatial data types

```
@Column(name="the_geom")
@Type(type="org.hibernatespatial.GeometryUserType")
private Point geometry;
```

The presentation layer (graphical user interface) is implemented with JSF technology.

- *WMF/WFS Server:* We use GeoServer[10], an Open Source server written in Java, which is the reference implementation of the Open Geospatial Consortium WFS, WMS and WCS specification. Since the required protocols WFS and WMS are standardized, every OGC

[8]Raster images or vector formats like Scalable Vector Graphics (SVG) and Web Computer Graphics Metafile (WebCGM).

[9]The XML Linking Language (XLink) is used to create internal and external links within XML documents.

[10]http://www.geoserver.org

compliant server like the UMN MapServer[11] or Degree[12] can be used.

- *Spatial Database:* As a spatial database PostgreSQL in conjunction with *PostGIS*, an open source addition to PostgreSQL to make the database capable of using geographic objects, is used. In 2006, PostGIS was certified as a compliant "Simple Features for SQL" database by the Open Geospatial Consortium.
  The geographic objects represented by *HibernateSpatial* are stored in the *Well-Known Text* (WKT) format. Compared to the standard PostgreSQL database engine additional geometric functions (like spatial joins, intersections etc.) are added to the database as well as spatial reference systems describing the geodetic datum, geoid, coordinate system and map projection of the spatial objects.

Beside the client side components (JavaScript) and the JSF component library itself, *GeoJSF* consists of several classes supporting the managed beans on the application server side. This includes event handlers for client-side events as well as data structures representing all map related objects. Client events with spatial information like point queries (a click on the map with a individual search radius) or the selection of an area with a traverse are forwarded to customizeable factory classes building a WFS query which is send to the WFS server. The result is interpreted and the corresponding entity objects (Entity EJB) are instantiated. While these objects are under control of an `EntityManager`[13] the complete object tree is available with all (sub) child elements without any additional effort for programmers, the elements are loaded on demand from the database.

B. User Interaction

The user interaction and integration of WebGIS components into a JSF application can be easily explained in the following example in Figure 4. The screenshot is taken from a geographic information system (with integrated water resource management) for the Lake Chad Basin Commission (LCBC) in West Central Africa funded by the the German Society for Technical Cooperation (GTZ).

The page contains different elements, the menu bar on the top and both panels on the right side are standard JSF components of the component library *RichFaces*, the chart at the bottom is generated by *JFreeChart*. The map on the left side as well as the overview map are generated by *GeoJSF* allowing interactions to and from other JSF components.

1) The user can select/deselect different map layers using the *GeoJSF* Layer-Control, the available layers are defined for each thematic map and legend symbols are generated on the fly by the Styled Layer Descriptor (SLD) of the actual layer.
2) Changes in the layer selector will immediately re-render the interactive map on the left side. The map is based on *OpenLayers* to allow scrolling and zooming and is enhanced with controls to interact with the JSF backend at the server.
3) Movements in the map component or selections in the overview map are directly reflected on the corresponding map.
4) A click on the map invokes the JSF backing bean and a WFS query is created with the actual coordinates and search radius. In this example a list of all rainfall stations is retrieved, the corresponding EJB entity beans are instantiated and the table is populated with the station name.
5) Optionally the stations to be included in the chart generation can be selected and deselected.
6) The chart for the measured rainfall of all selected stations in the specified time interval is shown in the chart panel at the bottom and immediatly updated by new map or station selections.

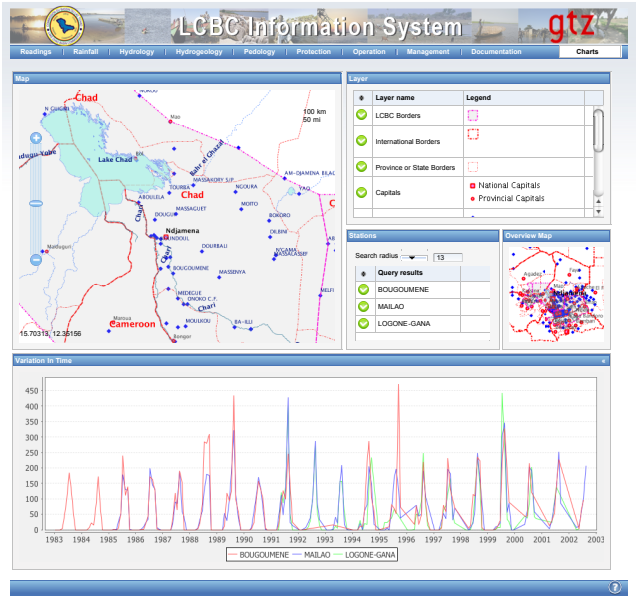


Figure 4. Example of a JSF application with *GeoJSF* maps

VI. Conclusion and Future Work

The JavaEE platform is a feature rich framework fulfilling all requirements to build up robust and scalable enterprise applications. During our developments with this technology on various information systems, ranging from water information systems, management information systems or systems for enterprise resource planning, we were faced with the client's demand on processing and visualization of spatial data. As a matter of course we did not want to change the

[11] http://mapserver.org
[12] http://deegree.org
[13] The `EntityManager` is part of the Java Persistence API (JPA) and located in the application server.

 ISBN: 978-1-61208-118-2

technology or switch to commercial solutions. Realizing that at this time no free available product was available satisfying our needs, we decided to develop *GeoJSF* in the beginning of 2010.

Our experiences show that *GeoJSF* easily enables Java EE programmers to include interactive maps in their JSF applications. The separated development domains for Java programmers dealing with Java EE on one side and GIS experts on the other side dealing with spatial information can easily consolidated with a XML based definition for thematic maps. These maps usually contain of different layer (groups) and can be customized by SLD totally independent of the Java development.

Currently we are preparing to publish the library under an Open Source license in the SourceForge software repository. This includes a considerable effort of documentation and usage examples. Other objectives are a more comprehensive interaction between map elements and JSF backing beans, e.g. markers or thematic overlay elements.

REFERENCES

[1] C. Granell, L. Diaz, and M. Gould, "Geospatial web service integration and mashups for water resource applications," in *Proceedings of ISPRS Congress Beijing 2008*, vol. XXXVI-I/B4/IV. International Society of Photogrammetry and Remote Sensing, 2008.

[2] J. Y. C.T. Yang, L. Ye, "A framework of web-gis for urban drainage network system," in *International Conference on Semantic Web & Web Services*, 2006.

[3] A. Sayar, M. Pierce, and G. Fox, "Integrating ajax approach into gis visualization web services," in *Proceedings of IEEE International Conference on Internet and Web Applications and Services ICIW'06 February 23-25, 2006*, 2006.

[4] S. A. Romalewski, "Rich interactive mapping experience through open source frameworks frameworks and ajax data visualization techniques," in *Proceedings of the ISPRS working group III/4, IV/8, IV/5*, T. H. Kolbe, H. Zhang, and S. Zlatanova, Eds. International Society of Photogrammetry and Remote Sensing, 2008.

[5] J. She, Q. Chen, S. Pan, and X. Feng, "Monitoring land use of construction based on webgis with ria technology," *Information Science and Engineering, International Conference on*, pp. 2104–2108, 2009.

[6] P. S. Talegaonkar, "Service oriented architecture for gis applications," in *The 12 International Conference of International Association for Computer Methods and Advances in Geomechanics (IACMAG)*, 2008.

[7] A. M. S. Fabiana Soares Santana, "Soc & soa in ecological niche modelling and agribusiness: Discussion and case studies," in *7th World Congress on Computers in Agriculture Conference Proceedings*, 2009.

[8] P. Townend, J. Xu, M. Birkin, A. Turner, and B. Wu, "Modelling and simulation for e-social science through the use of service-orientation and web 2.0 technologies," in *Proceedings of the 4th International Conference on e-Social Science*, 2008.

[9] J. Fleischer, R. Häner, S. Herrnkind, A. Kloth, U. Kriegel, H. Schwarting, and J. Wächter, "An integration platform for heterogeneous sensor systems in gitews: Tsunami service bus," *Natural Hazards and Earth System Sciences*, vol. 10, pp. 1239–1252, Jun. 2010.

[10] J. She, S. Pan, Q. Chen, X. Feng, H. Jiang, and K. Xiao, "Web-based integrative presentation of distributed spatial data," *Information Science and Engineering, International Conference on*, pp. 2233–2237, 2009.

[11] H. Lu, W. Nihong, W. Chang, and C. Yujia, "The research on the webgis application based on the j2ee framework and arcgis server," *Intelligent Computation Technology and Automation, International Conference on*, vol. 3, pp. 942–945, 2010.

[12] J. Wang, T. C. Stein, T. Heet, D. M. Scholes, R. E. Arvidson, and V. Heil-Chapdelaine, "A webgis for apollo analyst's notebook," *International Conference on Advanced Geographic Information Systems, Applications, and Services*, pp. 88–92, 2010.

[13] W. Jing-zhong and L. Hui-dan, "Research on the web gis technology based on mapxtreme," in *CSSE '08: Proceedings of the 2008 International Conference on Computer Science and Software Engineering*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 123–126.

[14] C. Granell, C. Abargues, L. Diaz, and J. Huerta, "Interlinking geoprocessing services," *International Conference on Advanced Geographic Information Systems, Applications, and Services*, pp. 99–104, 2010.

[15] M. Hasegawa, S. Bhalla, and T. Izumita, "A user oriented query interface for web-based geographic information systems," *Frontier of Computer Science and Technology, Japan-China Joint Workshop on*, vol. 0, pp. 106–113, 2007.

[16] "Building web-based spatial information solutions around open specifications and open source software," vol. 7, no. 4, pp. 447–466, 2003. [Online]. Available: http://doi.wiley.com/10.1111/1467-9671.00158

[17] G. H. Robert Berry, Richard Fry and S. Orford, "Bulding a geo-portal for enhancing collaborative socio-economic research in wales using open-source technonolgy," *Journal of Applied Research in Higher Education*, vol. 2, no. 1, pp. 77–92, January 2010.

[18] JSR-127 JavaServer Faces. [Online]. Available: http://jcp.org/en/jsr/detail?id=127

[19] Open Geospatial Consortium, Inc. (2010) Welcome to the ogc website. [Online]. Available: http://www.opengeospatial.org

[20] J. de la Beaujardiere, Ed., *OpenGIS Web Map Server Implementation Specification (1.3.0)*. Open Geospatial Consortium Inc., 2006.

[21] P. A. Vretanos, Ed., *Web Feature Service Implementation Specification (1.1.0)*. Open Geospatial Consortium Inc., 2005.

 ISBN: 978-1-61208-118-2

City Energy Management: A Case Study on the Urban Area of Liège in Belgium

Sigrid Reiter, Véronique Wallemacq
University of Liège
Local Environment Management and Analysis (LEMA)
Liège, Belgium
e-mail: Sigrid.Reiter@ulg.ac.be, vwallemacq@ulg.ac.be

***Abstract*— Within the framework of sustainable development, it is important to take into account environmental aspects of urban areas related to their energy use. In this research, a typology of urban blocks is drawn up for the urban area of Liège through the use of GIS tools in order to assess energy uses of residential buildings and transport of residents at the city scale. For each class of this typology, a representative block is selected in order to model energy use at the city scale, as well as to consider the possible evolution of the city energy consumption and to simulate the effects of some strategies of urban renewal. An application study on the residential buildings energy consumption part of this typology is given to compare different energy management strategies. This case study allow to conclude that the European Directive on the Energy Performance of Buildings and even more selective energy policies on new buildings are not sufficient to widely decrease the energy consumption of Liège building stock but that renovation of the existing building stock has a much larger positive impact on city energy consumption reductions. These conclusions put forward the benefits of using urban GIS for policymaking and city management. Energy management is an important GIS application field.**

Keywords - Urban GIS, energy consumptio, forecast scenarios.

I. INTRODUCTION

In the actual context of growing interests in environmental issues, reducing energy consumptions in the building and the transport sectors (which represent respectively 37% and 32% of final energy in the European Union) appears as important policy targets. Urban areas are supposed to present high potentialities in terms of energy reduction. However, existing models often adopt the perspective of the individual building as an autonomous entity, and neglect the importance of phenomena linked to larger scales [1].

This research focuses on city level energy management. In this paper, a typology of urban blocks is drawn up for the urban area of Liege (in Belgium) through the use of GIS tools in order to assess energy uses at the city scale. This typology of urban blocks is organized into two parts: the residential buildings energy consumption and the transport energy consumption of residents. For each topic of this typology, representative blocks are selected in order to model the energy use at the city scale, as well as to consider the possible evolution of the city energy consumption and to simulate the effects of some strategies of urban renewal. An application study uses this typology to compare the effects of the European Directive on Energy Performance of Buildings with even more selective energy policies on new buildings and with renovation strategies on the existing building stock of the urban area of Liège.

The structure of this paper is developed in eight sections: the introduction, the state of the art and method, the study area and chosen criteria, the cartographic work, the typology of urban blocks generated, the calculations of the energy performance of the city, a discussion on the results and the conclusion.

II. STATE OF THE ART AND METHOD

This section describes the most important references on city energy management and the methodology used in this research.

A. State of the art

There are a lot of modeling tools to assess energy management of a specific building. However, such an approach makes it difficult to generalize the results in order to determine the best strategies at the urban scale. On the other hand, there are two types of modeling methods used to predict energy consumption at a large scale (for example, for national predictions): the top-down and bottom-up approaches. These methodologies have already been described in details [2,3]. The top-down modeling is generally used to investigate the inter-relationships between the energy and economy sectors. They study the influence of economic variables such as income or fuel prices on the energy consumption of countries. These models lack details on the building stock to be able to quantify the effectiveness of some specific energy policy measures on the urban energy performance. Bottom-up methods are based on typologies and components clustering modeling approach. These components can be buildings [4,5], urban blocks or neighborhoods [6]. This implies that they need extensive databases to support the choice and description of each component of their typologies. This is usually done by a combination of building physics modeling, empirical data (for example from housing surveys), statistics on national or

 ISBN: 978-1-61208-118-2

regional data sets and some assumptions about buildings performance. The bottom-up method is very useful to assess the energy consumption of existing building stocks.

B. *The method*

This research uses an Urban GIS in order to develop an energy model of the residential building stock of Liège and to spatialize its major components. Our approach combines global statistics, that are not associated with buildings (top-down approach), with features related to buildings and urban form (bottom-up approach). The evolution of the number of buildings in the residential stock is deducted from global trends of recent years (top-down approach), while the energy consumption of buildings are obtained thanks to empirical data and results of buildings energy modeling (bottom-up approach). This combined approach provides a set of data as accurate as possible.

III. STUDY AREA AND CHOSEN CRITERIA

Our study is focused on the urban area of Liège and more specifically on its residential urban blocks. Delimitation of city blocks was performed using data from the PICC, that is a computer project of continued mapping from the Public Service of the Walloon Region of Belgium. These data are provided in the form of vector map layers that characterize the natural environment (rivers, forests), the built environment (buildings) and the infrastructure (roads, railways, etc.) at scale 1/1000. The spatial position of these objects is known by their position (x, y) and their altitude (z) with an accuracy of 25 cm.

The first part of this research develops a typology of Liège's urban blocks. First, a large number of variables were selected to characterize the energy efficiency of city blocks, using an extensive literature review on this subject. Then, a statistical treatment of these parameters was performed using a Principal Component Analysis. This methodology [7,8], allows crossing a large number of criteria and grouping them according to their similarities. This statistical treatment reduced the number of our selected criteria to characterize the energy performance of the residential building stock of Liège. These are the six chosen criteria:

- Buildings' date of construction (before 1930, from 1931 to 1969, from 1970 to 1985, from 1985 to 1996, from 1996 to today), depending on the types of construction related to Belgian regulations. These data are defined across the urban blocks from the cadastre.
- Type of buildings (two, three or four frontages). Indeed, a terraced house uses less energy than a separate house [9]. These data are defined across the urban blocks from the cadastre.
- Type of housing (collective or individual). These data are defined across the urban blocks from the cadastre.
- Urban functions (residential, trade, school or socio-cultural facilities, services). Each block may contain one to four of these functions. The functional mix reduces energy consumption associated with shorter distances between the different activities' locations of everyday life.
- Index of energy performance for residents' travels to their work places [10,11]. This index is based on statistical data available at the census block scale (that is the smallest geographical unit in which data are available in Belgium). These data come from national censuses, carried out every ten years in Belgium. Weighted average of the built area of each city block has been achieved to adapt these data from the census block scale to the scale of the city block.
- Potential modal shares for alternatives to the car, following [12]. This calculation takes into account the daily frequency of trains and buses weighted by their type and destination.

IV. CARTOGRAPHIC WORK

The information in the cadastre has been linked to the file map showing the layout of the plots using their cadastral number. Then, the spatial relationship between the plots and the PICC data was established through the ArcMap function "Spatial join". This helped to know the date of buildings construction given on the cadastral maps.

It is important to note that some plots of the PICC found no match in the database of the cadastre. No data will be taken into account for the buildings constructed on these plots. Note that these differences arise because the data from the PICC were developed from aerial rectified photographs and the data from the cadastre were developed from digital cadastral maps. However, these data can be considered acceptable because only 383 buildings could not be taken into account, which represents only 0.2% of the residential building stock of the urban area of Liège.

To convert, as consistently as possible, the data known at the census block to the scale of the urban block, the data associated with each statistical area were distributed in a grid, which has a resolution of 10 m wide (see Figure 1).

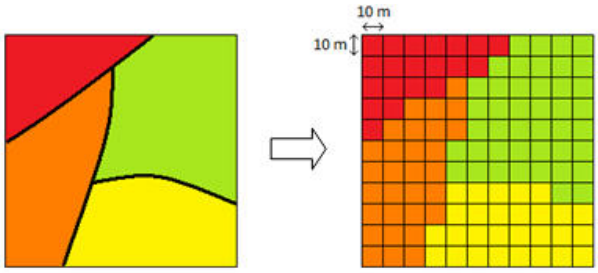


Figure 1 : Distribution of the data associated with four census blocks to a spatial grid of 10m wide.

Then, a weighting is applied according to the surfaces of the urban blocks that are related to one or several census blocks. For example, in Figure 2, the urban block value will

 ISBN: 978-1-61208-118-2

be calculated by adding twice the value of red cells, sixteen times the value of green cells, six times the value of yellow cells and 9 times the value of orange cells, and dividing the sum by 33 (the number of census cells covered by the urban block).

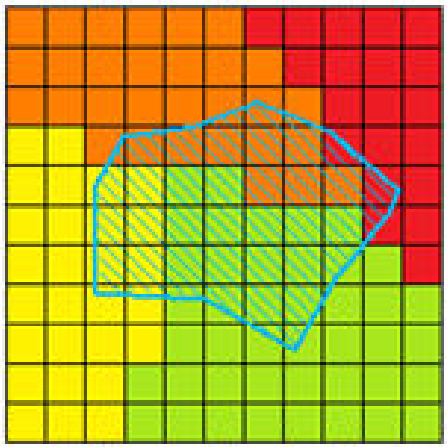


Figure 2: Wheighting calculation of the urban block (in blue) data on basis of four census blocks data.

However, it should be noted that the transport consumption criteria, based on this mapping work, are less accurate than the buildings consumption criteria, based on the cadastral values, because of the assumption that statistical data are evenly distributed in each census block.

V. TYPOLOGY OF URBAN BLOCKS

The typology of urban blocks is organized into two topics: residential buildings energy consumption and transport energy consumption of residents. For each topic of this typology, representative blocks are selected in order to model energy use at the city scale, as well as to consider the possible evolution of the city energy consumptions and to simulate the effects of some strategies of urban renewal. The proposed representative blocks were selected by choosing for each topic the representative criteria into the list of six criteria previously determined and then crossing them at the urban block scale. The six criteria that were taken into account are buildings date of construction, type of buildings (two, three or four frontages), type of housing (collective or individual), index of energy performance for residents' travels to their work places, potential modal shares for alternatives to the car and mix of urban functions (residential, trade, school or socio-cultural facilities, services).

For the first topic "residential buildings energy consumption", the following criteria were selected: buildings date of construction , type of buildings (number of frontages) and type of housing (collective or individual). After crossing these three criteria at the urban block scale, only the main classes, that include the largest number of urban blocks were selected. So, fourteen types of urban blocks have been defined, which represent 97% of the blocks of the urban area of Liège. For example, there are 508 blocks built before 1930, where over 66.6% of the buildings are terraced houses and most of them are individual housing; this type of urban block corresponds to 12% of the residential building stock of Liège (see Figure 3). Another block type is constructed after 1970, where over 66.6% of the buildings are separate and most of them are individual housings; there are 314 urban blocks of this type in the urban area of Liège, which corresponds to 7% of the residential building stock (see Figure 4). Within each of these 14 types, a representative block was chosen to allow modeling more accurately the energy consumption of buildings.

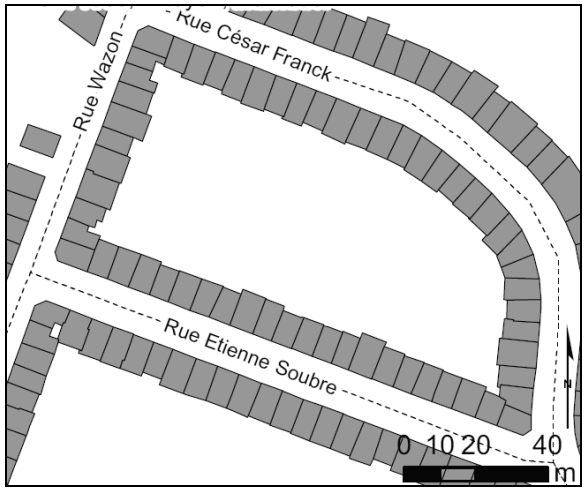


Figure 3: One type of urban block that represents 12% of the residential building stock of Liège.

Figure 4: One type of urban block that represents 7% of the residential building stock of Liège.

For the second topic "energy consumption by transport of residents", the following criteria were selected: the mix of urban functions (residential, trade, school or socio-cultural facilities, services), the index of energy performance for residents' travels to their work places and the potential modal shares for alternatives to the car.

 ISBN: 978-1-61208-118-2

After crossing these three criteria at the urban block scale, only the main classes, that include more than 0,5% of all the urban blocks, were selected. The selection of blocks representative of each class of theme 2 is performed identically to theme1.

VI. MODELING THE ENERGY PERFORMANCE

The three criteria that are taken into account to achieve the energy assessment of the residential housing stock are the age of buildings, their number of frontages and the type of housing (individual or collective). The energy consumption for each type of housing in the Walloon Region (including heating, hot water and lighting) are determined on basis of empirical values and simulation results. When these values are related to each building, it is possible to establish the evolution of the energy consumption of the whole urban area of Liège since 1850, that is the first date of construction of a building identified in the cadastre (see Figure 5). Before 1931, the dates of buildings construction are aggregated for periods lasting from 20 to 25 years, which explains the larger width of the bars in the Figure 5.

The most important actual energy policy measure in the EU is the Directive on the Energy Performance of Buildings (Directive 2002/91/EC) that came into force in 2002 with legislation in member states by 2006 [13]. These policy measures focus on energy efficiency when new buildings are constructed or when big buildings (larger than 1000m²) undergo a major renovation. However, there might be energy efficient measures that are environmental efficient and cost effective also on the existing residential building stock, on smaller buildings and/or lighter renovation processes. Note that in the Danish implementation of the EPB directive all existing buildings (including single family houses) are covered by the energy efficiency measures when they undergo a major renovation [4].

It is thus useful to model some forecast scenarios to compare the effects of the European Directive on Energy Performance of Buildings (EPB directive) with even more selective energy policies on new buildings and with renovation strategies on the existing building stock.

The demographic data of the population of our study area are known at the census block scale. The simplest hypothesis would estimate that the residential building stock changes proportionally to the population. However, the number of buildings in urban area of Liege during the last eight years did not increase as rapidly as the population during those years. We have thus established a base curve of the evolution of the built stock according to the statistics of its evolution between 2000 and 2008. This trend is represented by the following equation:

$$Y = 477{,}35 \ln(x) + 161\,348 \quad (1)$$

with x = forecast year – 2000 and Y = Number of buildings. This curve follows very well the recent trend of development of the residential building stock since the coefficient of determination calculated from the data observed between 2000 and 2008 amounts to 99.7%.

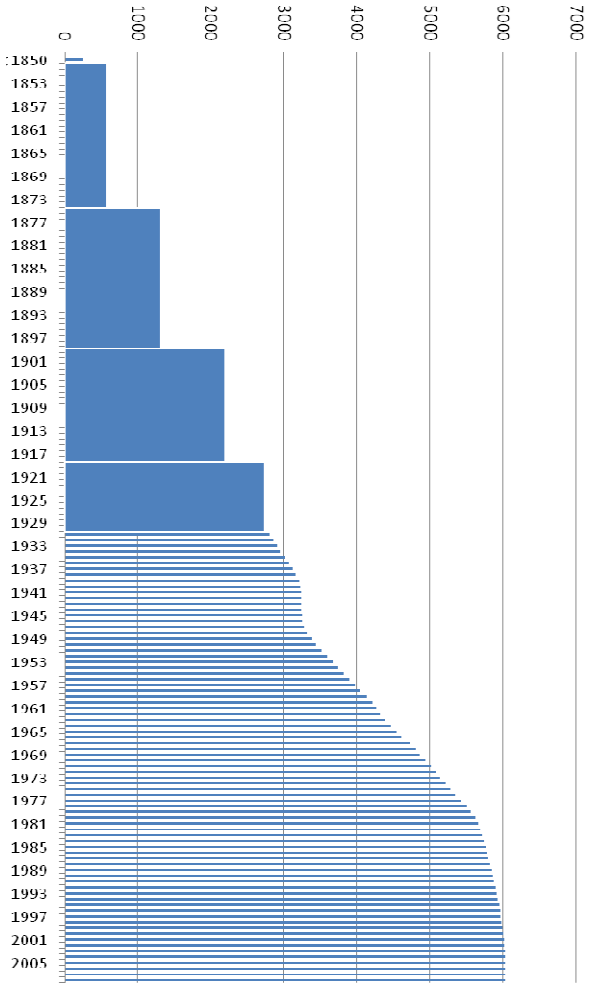


Figure 5: Evolution of the energy consumption of the urban area of Liège (in GWh/year) since 1850.

A. *Scenario 1: new buildings following EPB*

In this first scenario, the existing building stock remains unchanged, but new buildings are constructed according to the actual standard on the energy performance of buildings (EPB): the building's energy consumption should not exceed 115 kWh/m² per year. It is therefore the most likely evolution of Liège's building stock if the energy policies are not changed in the future. Following this first scenario, the energy consumption for the city of Liège in 2061 is estimated at 6067.74 GWh per year.

B. *Scenario 2: strengthening of energy policy on new buildings*

Considering that 5% of new housing stock will have low energy performances (LE: 95 kWh/m² per year), 2% of

buildings very low energy performances (VLE: 65 kWh/m² per year) and 1% will reach the standard passive house (50 kWh/m² per year), energy consumption decreases from 679 MWh for the year 2061 compared to the first scenario, which represents a reduction of only 0.01% for a period of fifty years.

Achieving 10% reduction in energy consumption of all buildings constructed after 2010 would require that the new stock meets the following constructive standards: 63% of buildings achieving the EPB standard, 21% of LE buildings, 10% of VLE buildings and 5% of passive buildings. But on the whole building stock, this reduction generates a very small decrease in energy consumption (0.06%) compared to Scenario 1, corresponding to the actual regulations.

C. Scenario 3: roof insulation of the old building stock

Following Verbeek and Hens, insulation of the roof is the most effective and durable measure for energy performance increase of households in Belgium [14].

A rate of renovation of buildings of 0.6% per year is chosen to simulate a realistic policy for roof insulation of the existing building stock equal to two thirds of the total rate of renovations observed in the Walloon Region on an annual basis. It is also assumed that the energy management is carried out efficiently: the oldest and least energy efficient buildings are the first to be renovated. Renovating the roof insulation of this old building stock will be incorporated as a reduction of 40% of energy consumption in comparison to the initial energy performance of these renovated buildings.

It appears that the renovation of existing buildings can drastically reduce energy consumption across the urban area. The total estimated consumption amounts to 5439.27 GWh/year in 2061, of which 99.5% is attributed to the existing stock. The decrease in total energy consumption is therefore 10.36% (628.46 GWh/year) compared to 6067.74 GWh/year for Scenario 1.

D. Scenario 4: renovation of the old building stock reaching EPB

This scenario aims to assess the amount of energy that could be saved if the existing building stock was renovated, at a rate of 0,6% per year, to meet the current EPB standard in Belgium (115 kWh/m² per year), while all the new buildings meet the same energy performances. Following this scenario 4, the estimated energy consumption for the city of Liège reach 5307.20 GWh/year in 2061. It is 760.54 GWh/year (13%) less compared to scenario 1.

E. Scenario 5 : renovation of all the existing building stock reaching EPB

The renovation of all the buildings of the residential building stock of Liège to the level of the current EPB standard in Belgium (115 kWh / m² per year), would result in significant reductions in energy consumption of the urban area, see Figure.6. Indeed, the global energy consumption would drop to 3178.23 GWh/year only, which represents a reduction of 47.6% compared to 6067.74 GWh/year of the scenario 1 (where the new buildings reached already the standard EPB, but where no renovation was undertaken).

However, to achieve the complete renovation of the existing housing stock by 2060, the rate of renovation of the urban area of Liege should increase sharply, to a minimum of 1.92% per year, which would require strong policies to accelerate and strengthen the process of renovating existing buildings.

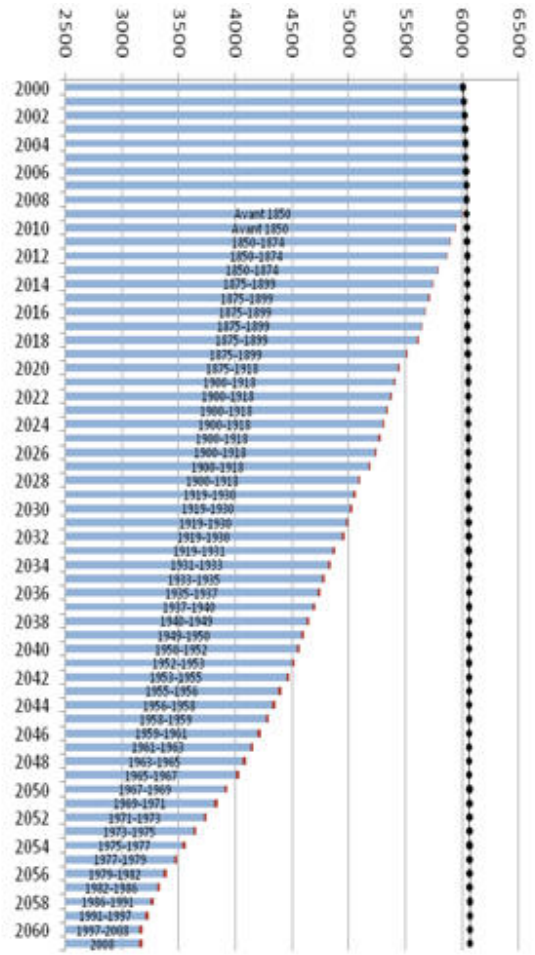


Figure 6: Energy consumption of the urban area of Liège (in GWh/year) from 2000 to 2061, following the scenario 5 (blue results) and comparison with scenario 1 (black dotted curve).

F. Scenario 6 : all the existing building stock reaching EPB and new buildings reaching the passive standard

This scenario uses the same renewal policy that the previous scenario but it is also assumed that each new housing built from 2012 will reach the passive standard (50 kWh/m² per year). The result of scenario 6 is very close to the previous scenario. The total energy of the urban area in 2061 amounts to 3161.57 GWh/year, which represents only a reduction of 0.5% compared to scenario 5.

 ISBN: 978-1-61208-118-2

VII. DISCUSSION

The studied scenarios show that the actual city energy challenge lies mainly in the renovation of the existing building stock. Indeed, the first two scenarios and the small difference between scenarios 5 and 6 show that it is not possible to ensure a significant reduction in energy consumption at the city scale applying only energy policies for new buildings, like the standard EPB already in use or by enhancing the performance of new buildings to low energy level, very low energy level and even to the passive housing standard.

However, scenarios of existing housing stock renewal (scenarios 3 to 5) can significantly reduce the overall consumption of the urban area of Liege in the following proportions:

- 10.36 % of energy consumption reduction in 2061 through the roof insulation of the oldest buildings at a renovation rate of 0.6% of the building stock per year.
- 13 % of energy consumption reduction in 2061 through a renovation reaching the EPB level of the oldest buildings at a renovation rate of 0.6% of the building stock per year.
- 47.6 % of energy consumption reduction in 2061 through a renovation reaching the EPB level of all the existing residential building stock, which corresponds to a renovation rate of 1.92 % per year.

Thus, the National climate change targets in Belgium will be impossible without a strategic increase of the existing housing stock renovation.

VIII. CONCLUSION

In this research, a typology of urban blocks is drawn up for the urban area of Liege through the use of GIS tools in order to assess energy uses of residential buildings and transport of residents at the city scale. An application study on the residential buildings energy consumption part of this typology is given to compare different energy management strategies. This case study allow to conclude that the European Directive on the Energy Performance of Buildings and even more selective energy policies on new buildings are not sufficient to widely decrease the energy consumption of Liège's building stock but that renovation of the existing building stock has a much larger positive impact on city energy consumption reductions. This research also proves the benefits of using urban GIS for city management and policymaking.

ACKNOWLEDGMENT

This research has been funded by the Special Funds for Research of the French Community of Belgium, within the University of Liege.

REFERENCES

[1] C. Ratti, N. Baker, and K. Steemers, "Energy consumption and urban texture", Energy and Buildings, vol. 37 (7), 2005, pp. 762-776.

[2] L.G. Swan and V.I. Ugursal, "Modeling of end-use energy consumption in the residential sector: a review of modeling techniques". Renewable and Sustainable Energy Reviews, vol. 13, 2009; pp. 1819-1835.

[3] M. Kavgic, A. Mavrogianni, D. Mumovic, A Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic, "A review of bottom-up building stock models for energy consumption in the residential sector", Building and Environment, vol. 45, 2010, pp. 1683-1697.

[4] H. Tommerup and S. Svendsen, "Energy savings in Danish residential building stock", Energy and Buildings, vol. 38, 2006, pp. 618-626.

[5] A. Uihlein and P. Eder, "Policy options towards an energy efficient residential building stock in the EU-27", Energy and Buildings, vol. 42, 2010, pp. 791-798.

[6] Y. Yamaguchi, Y. Shimoda, and M. Mizuno, "Proposal of a modeling approach considering urban form of evaluation of city level energy management", Energy and Buildings, vol. 39, 2007, pp. 580-592.

[7] L. Lebart, A. Morineau, and J.-P. Fenelon, Traitement des données statistiques – méthodes et programmes, 2d ed., Bordas Éd., Paris : Dunod, 1982.

[8] M. Volle, Analyse des données, 3rd ed., Paris: Economica, 1993.

[9] A.-F. Marique and S. Reiter, "A method to assess global energy requirements of suburban areas at the neighbourhood scale", Proc. IAQVEC 2010 Conference, Syracuse (USA), 2010.

[10] K. Boussauw and F. Witlox, "Introducing a commute-energy performance index for Flanders", Transportation Research Part A , vol. 43, 2009, pp. 580–591.

[11] A.F. Marique and S. Reiter, "A method to assess transport consumptions in suburban areas", Proc. PLUREL Conference : Managing the Urban Rural Interface, Copenhagen, 2010.

[12] Y. Cornet, D. Daxhelet, J.-M. Halleux, A.-C. Klinkenberg and J.-M. Lambotte, « Cartographie de l'accessibilité par les alternatives a la voiture », Conférence Permanente du Développement Territorial, Belgium, 2005.

[13] Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the Energy performance of buildings, Official Journal of the European Union L 001, 2003, pp. 65-71.

[14] G. Verbeek and H. Hens, "Energy savings in retrofitted dwellings : economically viable?"; Energy and Buildings, vol. 37, 2005, pp. 747-754.

 ISBN: 978-1-61208-118-2

Specification for a Shared Conceptual Layer in GIS

Enguerran Grandchamp
Département Mathématiques et Informatique
LAMIA, Université des Antilles et de la Guyane
Pointe-à-Pitre, Guadeloupe France
e-mail: egrandch@univ-ag.fr

***Abstract* - During the last decade, GIS have suffered of the lack of structured information and semantic definition of the data. The numerous origins of the data and the spatial localization of the object inferred many inconsistencies among them. Indeed, the way to collect data and the level of sampling of the boundary differ from one application to another and from one expert to another. This leads to different representations of the same reality and different models of semantically identical objects or boundaries. To allow the matching of the objects, or parts of the objects, when combining different information layers we introduced a conceptual layer which described a scene with concepts and spatial relationships based on the definition of an ontology. This layer allows detecting and solving conflicts between the common boundaries of two objects. We present in this paper the specification and the way to build and use the ontology.**

Keywords - GIS; Information Fusion; Ontology.

I. INTRODUCTION

Nowadays, geographic information invades our life with more and more applications (navigation, localization, survey, security, etc for tourism, energy, environment, urban deployment, etc.). This growing of applications requires more and more data collected by different ways and different peoples leading to a great heterogeneity.

Geographic database are not well structured and consistent because of the numerous and non experts producers of information. Indeed, there are so many information layers and terminologies as people that produce them.

Moreover, the way to collect the data could be very different from high technologic tools (GPS receptors) to visual lecture on a map. Depending on the application, the expert needs few descriptions of the objects geometry or on contrary detailed ones.

All these considerations lead to different representations of the same reality with different accuracy (localization, description, etc.). This is particularly the case for natural or semi-natural limits such as forest, agriculture, etc. which are not well defined or localized (in contrast to buildings for example). The main problem is that most of the representations do not coincide; but we could not say that one is worst or better than the others. This is just an arbitrary choice of the vertices to sample the objects.

If the impact of such differences is not a problem for a thematic exploitation, there is a consequently propagation of errors when dealing with multi thematic problematic.

Indeed, the main drawback of this heterogeneity is that the power of the spatial analysis is reached when combining different thematic layers (agriculture and environment, roads and emergency, rainfall and population, etc.) and due to localization mistakes this combination leads to inconsistent information. Applications like the ones presented in [13] and [14] suffer from these localizations errors.

In order to avoid mistakes, we decide to refer to a conceptual view of the scene which describes the objects in terms of spatial relationships, geometry and semantic.

This view is an ideal representation of the scene and is based on the definition of an ontology and spatial relations.

An ontology is considered as being a set of concepts and relation between them. It gives semantic information to the scene.

Section 2 presents related works on spatial ontologies, ontology integration in GIS and localization errors correction. Section 3 explains the way to build the ontology. Section 4 deals with the conceptual layer. Section 5 explains the way to use the conceptual layer to solve conflicts and then Section 6 gives the conclusions and perspectives of this study.

 ISBN: 978-1-61208-118-2

II. RELATED WORKS

We present here related works concerning ontology integration in GIS and spatial object localization correction.

As Fonseca et al. [7][8] and later Cruz et al. in 2005 [5], defined an ontology driven GIS, there isn't any operational framework for an ontology integration in GIS.

The association between spatial data and ontology is done with external tools such as SPIRIT [18] or GeoSVM [7].

The first step in integrating ontologies in GIS is to define ontologies on spatial data. The main objective of these ontologies is to represent concepts linked to a specific expert domain (biology, geology, tourism, urgency, etc.) and also spatial relationships between them [25]. Many authors ([7], [8], [16], [23]) propose specific ontologies which differs from the used terms and features to describe the concepts. Indeed, depending on the application and domain some concepts could be described by different ways and with different levels of detail as if they are semantically identical.

The choice of the most adapted ontology for an application is not easy.

The second step is to introduce ontologies in GIS. Many authors introduce them in different ways.

Viegas et al. [24] and Baglioni et al. [2] create an intermediate semantic layer between the user and the geodatabase in order to facilitate the user's queries.

In [6], [10] and [11] the spatial ontology is used to facilitate object classification.

Concerning the correction of localization errors of spatial objects, to the best of our knowledge, the only approach to solve the problem of corresponding points between different information layers is expressed in [26] and uses buffer area to fuse vertices. This approach has two main drawbacks: the process is applied over all vertices *(i)* without distinction between semantically different ones *(ii)* without integrating spatial relationships.

III. ONTOLOGY BUILDING

There are many ways to build a geographic ontology. Many different ontologies had been defined in the literature, each one done for a specific application.

We have to choose among them the most appropriate ones. The first thing to do is to extract the *Intersection Knowledge* between these ontologies in order to have a common and recognized basis. After that, we select the *Augmenting knowledge* linked to a specific domain or expert [12]. To compare these ontologies and detect similarities we have to define metrics ([9], [19], [21]).

Figure 1. extracted from [12] summarizes the notion of *Intersection and Augmenting Knowledge*.

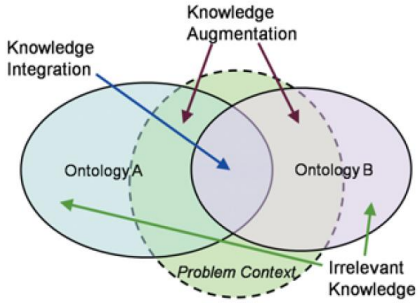


Figure 1. Ontologies intersection ([12])

We built the ontology by joining different independent ontologies compared to more general ontologies using similar or synonym terms ([17], [20], [24]) with the following criterion: *(i)* recognized ontology or commonly used one, *(ii)* built with a complete hierarchical object model, *(iii)* written with OWL (Ontology Web Language). In particular, we use SWEET (Semantic Web for Earth and Environmental Terminologies) ontology.

The ontology is introduced in the GIS using annotations in the data and in the metadata [22].

There is also automatic process to produce the ontology from the geographic data themselves [1], but the restricted area of interest and the numerous experts domains limits the advantage and efficiency of such a method and we prefer a manual building based on metrics evaluation.

IV. THE CONCEPTUAL LAYER

The process presented in this article is summarized in figure 2. Objects A and B come from different layers and represent different kind of objects or not. Depending on the application we decide to trust the boundary of object A (the case illustrated) or B. Decision criterions could be *(i)* the scale sampling of the objects *(ii)* the distance in the semantic space between the belonging class of A and B and the application (i.e. we prefer to trust river boundaries in an application concerning water). In the illustrated example, the boundary between A and B must be the same in the conceptual layer (example: A and B are to related agricultural fields).

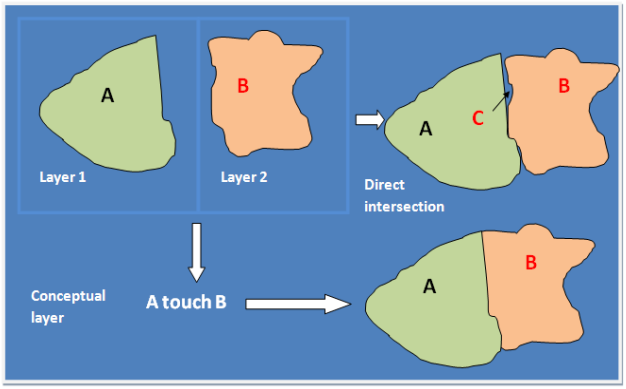


Figure 2. Process

 ISBN: 978-1-61208-118-2

But the boundaries are not the same in the two layers (Area C appears between objects A and B). We decide to conserve the boundary of object A because it is the most reliable one in the context of the application.

Because they will always have different ways to model the same object, we do not have to adopt a unique representation. As an example, depending on the observation scale, a building could be seen as a point, a square, or a more detailed polygon. Depending on the expert, the focus will be done on environment objects or urban ones. So we prefer adding semantic information concerning the data and the instances in order to have a common reference to compare the different view of the same data.

But nowadays, there isn't any way to define a vertex or a boundary as being an external resource or reference. As an example, if in a soil occupation layer we want to refer to a river defined in a hydrology layer we couldn't do it in a simple and normalized way. We can only do it during the layer creation step if we define an object as being the intersection or union of others object taken from sources layers. But as soon as the layer is created there is no link kept between layers and objects. So if we update the sources layers, changes are not propagated to the built layer.

To solve this problem we propose a common conceptual layer which regroups an ontology, representing the concepts of the scene, and the spatial relationships (topology) between concepts and instances.

A. Ontology

The ontology regroups the concepts present in the scene. We have to link these concepts with each object in the information layer in order to specify that an object is an instance of one or several concepts. To do so, we add a descriptive field in the table of the information layer which refer to concepts in the ontology and give information such as *is instance of*.

B. Spatial Relation

The proposed conceptual layer is a kind of extension in the GIS environment of the spatial relation hierarchy proposed in [4].

The authors of [4] define three levels for the representation of spatial relationships: the geometric level, the computer level and the user level. The geometric level is an abstract representation of the relations that we can consider to be exact (we use this level for the conceptual layer). The computer level takes into account the object reality and is a spatial relationships and geometric view of the information layer. If object geometry or relationships are altered in the information layer, this level will also be altered. We will compute this level directly from the data by analyzing the spatial relation between objects in order to qualify and quantify the relations. The user level is a restricted view of the information linked to the domain of the user and won't be used in the scope of this framework.

At the geometric level, we characterize: *(i)* the relations between the concepts of the ontology (more a qualification than a quantification). For example, the concept *private house* could be considered as being *linked to* the concept *road*. *(ii)* the relations between the instances of the concept (more a quantification than a qualification). For example, *house n°123, touch, road n°12*.

To characterize the spatial relations, Egenhofer [27] defines topological relations with intersection and overlapping (based on 4 or 9 intersection). These relations take into account:

1. The relative position which are strict (touch, intersects, etc.) or vagueness (close to, etc.). They could be binary (on the left, far from, etc.) or ternary (between [3]), and rarely quaternary (neighborhood).
2. The proportion (partially, completely, etc.) which are also vagueness.
3. Uncertainty: perhaps, certainly.

Some of these relations imply dependences between objects [15]. As an example *Objects A and B are around C* doesn't implies constraints between A and B but in the formulation *Objects A and B are on both sides of C* implies a dependences between the location of A and B.

Among all these characterizations of the spatial relationships, the strict relative position such as *touch*, *intersect*, *within or disjoined* are useful to solve conflicts. Other relations such as *close to* or *perhaps* don't give enough information to solve conflicts.

V. SOLVING CONFLICTS

Now let us look at the way to use the conceptual layer to solve conflicts between different representations of the same topological elements (objects or boundaries).

Firstly, it's easy to compute the computer level. We only have to describe each spatial relation between objects of the scene. Secondly, it's easy to compare the computer level with the geometric level and to localize the conflicts (as an example Object A intersects Object B in the computer level and Object A touches Object B in the geometric level). This could be done with the concepts or instances. Because most of the objects are localized with a relative good accuracy, conflicts between the two levels will not concerned relations such as *far from* instead of *close to*. Most of the conflicts will be generated by close objects having touching or merged boundaries in the geometry level and intersected one in the computer level.

Thirdly, we have to solve conflicts leading to unwanted intersections when combining layers. The aim is not to

 ISBN: 978-1-61208-118-2

change the representation of the objects and to modify the layers but only to intervene during the combination (union or intersection) of the layers. So we built the results of the combination by considering the geometry level and by choosing a unique representation for the objects concerned by the conflict. The choice is done by considering the domain and accuracy of each layer. The priority is given to the layer close to the domain of the expert using the GIS or to the more accurate one. Indeed, as explained earlier in this paper, the same object or boundary could be described at different scales leading to matching problems.

VI. CONCLUSION AND FUTURE WORK

This paper present the specifications for a conceptual layer shared between information layers in a GIS. The main objective of this layer is to propose an intermediate step to solve the matching problem of two corresponding objects or boundaries during the combination of different information layers. Indeed, if two vertices representing the same topological or semantic objects don't match, the union or intersection of the layers will lead to inexistent objects.
By referring to the conceptual layer, which regroups both concepts and spatial relationships, we will fuse the two different representations of the same object and choose a unique representation according to the application.

At present, the proposed framework is only at the specification step, but every technological ways to implement it have been studied and the implementation has already started and will be presented in a future paper.

REFERENCES

[1] F. Fonseca, M. A. Rodríguez, and S. Levashkin, "Building geospatial ontologies from geographical databases", In Proceedings of the International Conference on GeoSpatial Semantics (GeoS 2007), Springer Lecture Notes in Computer Science, vol. 4853, pp. 195–209, Berlin, 2007

[2] M. Baglioni et al., "Ontology-supported Querying of Geographical Databases", In Transactions in GIS, vol. 12, issue s1,_pp. 31–44, december 2008

[3] I. Bloch et al., "Modeling the spatial relation between for objects with very different spatial extensions", In Proceedings of Reconnaissance des Formes et Intelligence Artificielle (RFIA'06), Tours, France, Janvier 2006

[4] E. Clementini and R. Laurini, "Un cadre conceptuel pour modéliser les relations spatiales", In Revue des Nouvelles Technologies de l'Information (RNTI, vol. 14, pp. 1-18, novembre 2008

[5] I. Cruz, W. Sunna, and K. Ayloo, "Concept-level matching of geospatial ontologies", In Proceedings of GISPlanet, Lisbon, Portugal, 2005

[6] S. Derivaux, G. Forestier, C. Wemmert et S.Lefèvre, "Extraction de détecteurs d'objets urbains à partir d'une ontologie", In proceedings of Extraction et Gestion de Connaissances (EGC), vol. 2, pp. 71-81, Sophia, 2008

[7] F. Fonseca, M. Egenhofer, P. Agouris, and G. Câmara, "Using ontologies for integrated geographic information systems", In Transactions in GIS, vol. 6, pp. 231–57, 2002

[8] F. Fonseca, J. Martin, and A. Rodríguez, "From geo to eco-ontologies", In Proceedings of the International Conference on Geographic Information Science, pp. 93–107, Boulder, Colorado, 2002

[9] F. Fonseca, G. Camara, and A. M. Monteiro, "A framework for measuring the interoperability of geo-ontologies", In Spatial Cognition and Computation, vol. 6, pp. 309–31, 2006

[10] G. Forestier, S. Derivaux, C. Wemmert et P. Gançarski, "Interprétation d'images basée sur une approche évolutive guidée par une ontologie", In proceedings of Extraction et Gestion de Connaissances (EGC), vol. 2, pp. 469-474, Sophia, 2008

[11] G. Forestier, S. Derivaux, C. Wemmert et P. Gançarski, "An Evolutionary Approach for Ontology Driven Image Interpretation", In Proceedings of the International Conference on Applications of Evolutionary Computation (Evo'08), Springer, Lecture Notes in Computer Sciences, vol. 4974, pp 295--304, Italy, 2008

[12] M. Gahegan et al., "A Platform for Visualizing and Experimenting with Measures of Semantic Similarity", In Ontologies and Concept Maps, Transactions in GIS, vol. 12 issue 6, pp. 713-732, 2008

[13] E. Grandchamp, "GIS information layer selection directed by remote sensing for ecological unit delineation", In IGARSS, 2009

[14] E. Grandchamp, "Raster-vector cooperation algorithm for GIS Application to ecological units delineation", In GeoProcessing, 2010

[15] H. W. Guesgen and J. Albrecht, "Imprecise reasoning in geographic information systems", In Fuzzy Sets and Systems, vol. 113, pp. 121-131, 2000

[16] L. Gutierrez, "WP 9: Case study eGovernment D9.9 GIS ontology, "http://dip.semanticweb.org/documents/D9-3-improved-eGovernment.pdf ", last access : 17/01/2011, 2006

[17] C. Hudelot, J. Atif, and I. Bloch, "Ontologie de relations spatiales floues pour le raisonnement spatial dans les images", In proceedings of Représentation et Raisonnement sur le Temps et l'Image (RTE), 2006

[18] Jones C. B. et al., "Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT project", In Special Interest Group on Information Retrieval (SIGIR), ACM Press, pp.387 – 388, Finland, 2002

[19] M. Kavouras, M. Kokla, and E. Tomai, "Comparing categories among geographic ontologies", In Computers and Geosciences, vol 31, pp 145–54, 2005

[20] N. E. Maillot and M. Thonnat, "Ontology based complex object recognition", In Image and Vision Computing, vol 26, pp 102–113, 2008

[21] M. A. Rodriguez and M. Egenhofer, "Determining semantic similarity among entity classes from different ontologies", In IEEE Transactions on Knowledge and Data Engineering, vol. 15, pp. 442–56, 2003

[22] N. Schuurman et al.,"Ontology-Based Metadata", In Transactions in GIS, vol. 10, pp. 709–726, 2006

[23] B. Smith, "Ontology and information systems", http://www.loa-cnr.it/Papers/FOIS98.pdf, last access : 17/01/2011, 2001

[24] R. Viegas and V. Soares, "Querying a Geographic Database using an Ontology-Based Methodology", In Brazilian Symposium on GeoInformatics (GEOINFO 2006), pp. 165-170, Brazil 2006

[25] N. Wiegand et al., "A Task-Based Ontology Approach to Automate Geospatial Data Retrieval", In Transactions in GIS, vol. 11, issue 3, pp. 355–376, 2007

[26] P. A. Zandbergen, "Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National standard for Spatial Data Accuracy", In Transactions in GIS, vol. 12, issue 1, pp. 103–130, 2008

[27] M. Egenhofer and R. D. Franzosa, "Point-Set Topological Spatial Relations", In International Journal of Geographical Information Systems, vol. 5(2), pp. 161-174, 1991

 ISBN: 978-1-61208-118-2

Understanding Geographic Routing in Vehicular Ad Hoc Networks

Reinaldo Bezerra Braga
Joseph Fourier University (UJF)
Grenoble Informatics Laboratory (LIG)
STEAMER Team
braga@imag.fr

Hervé Martin
Joseph Fourier University (UJF)
Grenoble Informatics Laboratory (LIG)
STEAMER Team
herve.martin@imag.fr

***Abstract*—Geographic routing in Vehicular Ad hoc Networks (VANETs) has recently received considerable attention. Developing multi-hop communication in VANETs is a challenging problem due to the rapidly changing topology and network disconnections. These problems result in failures or inefficiency in traditional routing protocols used in Mobile Ad hoc Networks (MANETs). With the widespread adoption of Global Position System (GPS) and the progress on self-configuring localization mechanisms, geographic routing protocols offer promising solutions for message delivery. In this paper, we present an overview of geographic routing strategies in vehicular ad hoc networks. In addition, we introduce the main challenges of using geographic routing protocols in VANETs from different perspectives and discuss some directions of future research on this field.**

***Keywords*-Geographic Routing Protocols; Geographic Information Systems; Vehicular Ad Hoc Networks**

I. INTRODUCTION

Mobile ad hoc networks (MANETs) are self-configuring and self-organizing multi-hop wireless networks, composed by a set of mobile nodes that move around the network and cooperate in transmitting packets among the nodes. A MANET performs efficient and robust procedures by providing routing functionalities for mobile nodes. For instance, a unicast routing creates a multi-hop forwarding path for a pair of source and destination nodes beyond the direct wireless communication range. In addition, routing protocols maintain connectivity if the links on the paths break due to some problem, such as radio propagation, node movement, or wireless interference.

In MANETs, the velocity of a mobile node is probably equal to that of a walking person. If mobile nodes are vehicles, these networks are called Vehicular Ad Hoc Networks (VANETs). Compared with MANETs, the velocity of vehicles in VANETs is much higher since vehicles move faster than walking persons [1]. The main motivation to study routing protocols in VANETs is related to the expansion of data exchange among vehicles in order to provide robust applications for Intelligent Transportation Systems (ITS). VANET applications can include on-board active safety systems, providing communications among nearby vehicles (V2V) and between vehicles and the roadside infrastructure (V2I). However, several challenges have been identified to adopt VANET utilization on a large scale. The challenges are usually associated with the high node mobility, dynamic scenarios and the scalability considering the number of nodes. Therefore, it is important to develop a robust routing protocol to provide an efficient communication among nodes.

Currently, there are mainly two types of routing protocols in VANETs: topological routing and geographic routing. In topological routing, mobile nodes use topological information to manage routing tables or search routes directly. In geographic routing, each node knows its own position and makes routing decisions based on the position of the destination and the positions of its local neighbors [2]. With the widespread adoption of Global Position System (GPS) and the progress on self-configuring localization mechanisms, geographic routing in VANET has garnered significant attention to provide promising solutions for message delivery.

In spite of the existence of a considerable number of papers about geographic routing in mobile ad hoc networks (MANETs) [3] [5], we perceived a lack of a specific overview involving the use of geographic routing protocols in vehicular ad hoc networks (VANETs). This paper then presents an overview of general concepts of geographic routing in vehicular ad hoc networks. In addition, we introduce the main challenges of using geographic routing protocols in VANETs from different perspectives and discuss some directions of future research on this field. The remainder of this paper is structured as follows. We first introduce basic concepts of geographic routing and present some general goals for designing a routing protocol in VANETs. We then present the geographic routing protocols available in the literature, followed by a discussion about these protocols. Finally, we indicate some possible directions of future research and conclude the paper.

II. BASIC CONCEPTS OF GEOGRAPHIC ROUTING

Although the research on geographic routing being more recent than topological routing, it has received a special attention due to the significant improvement that geographic information can produce in routing performance. Geographic routing can be defined as a type of stateless routing, in which it is not required that a node performs maintenance

 ISBN: 978-1-61208-118-2

functions for topological information beyond its one-hop neighborhood [6]. Consequently, geographic routing is more feasible for large-scale networks than topological routing, which requires network-wide control message dissemination. Besides that, geographic routing requires lower memory usage on nodes by maintaining the information locally.

In general, geographic routing is composed of two main components: a location service and a geographic forwarding process. The location service determines the position of the packet destination in order to improve the routing process for creating the path with source node, using intermediary nodes. Consequently, the position of the packet destination can be added in the packet header so that intermediate hops can know where the packet is destined for [7].

Likewise, geographic forwarding is performed in two modes, namely, geographic greedy-forwarding mode and void-handling mode[1]. The greedy-forwarding mode defines a next-hop node for packet forwarding taking into account the positions of the current node, its neighboring nodes, and the destination node. A node can obtain its own position via a GPS receiver or through other localization algorithms. The positions of the neighboring nodes can be acquired either from a centralized neighborhood table at the node or in a distributed method via contention among neighboring nodes [10]. At last, the position of the destination node is included in the packet header sent by the source node. However, if some intermediate node knows a more accurate position of the destination, it is able to update the position in the packet header before forwarding the packet.

Geographic routing protocols offer some advantages over traditional ad hoc routing strategies. First of all, the geographic forwarding process allows the path adaptation by selecting the best next hop, if an intermediate node, previously used, becomes unavailable. Due to the absence of a routing creation process, this path selection does not need a table maintenance procedure other than intermediate neighbors and the propagation of control packets. Other advantages are related to the capacity to utilize weight additional metrics in order to select the next hop and the route alteration node by node taking into account the QoS related to the neighbors, such as bandwidth and delay [11]. However, some challenges in geographic routing are still open and need to be investigated [12].

- The difficulty to control the overhead required for distributed location database service of geographic routing protocols. Although location based addressing offers a convenient, naturally occurring, hierarchical address structure in terms of name, city, state and country, it may lead to excessive control overhead in conditions of high mobility.
- The irregular distribution of vehicles on urban centers makes route selection more complex, e.g. the shortest path protocols may produce more frequent network disconnections.
- High signal interference in the communication due to the presence of large buildings. Therefore, a building or lack of radio coverage can result in voids in the physical network topology. These voids can obstruct the packet forwarding process at local minima, where the neighbors close to the destination are hidden/unreachable, resulting in a failure.

In summary, the process to compute the best routes to send packets in Vehicular Ad hoc Networks (VANETs) is a difficult task due to high node mobility and the existence of unstable wireless links. To improve the performance of geographic routing protocols, several solutions were created. We present an overview of the main proposed techniques for geographic routing and appoint the main challenges of using geographic routing in vehicular ad hoc networks.

III. Geographic Routing in Vehicular Ad hoc Networks

Ad hoc routing protocols have to provide routing procedures to select the best routes in order to send packets from a source node to a destination node, taking into account the utilization of multiple hops. In the same way, geographic routing protocols can use location services to improve these procedures. Figure 1 shows a general architecture of geographic routing in VANETs.

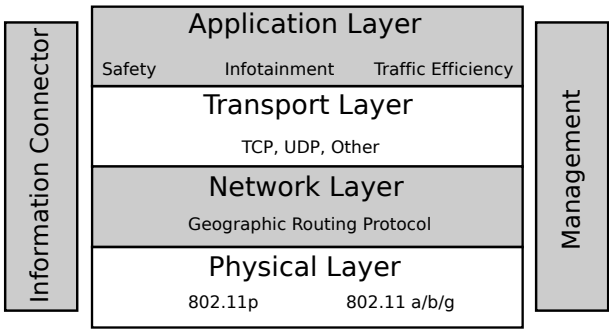


Figure 1. Geographic routing architecture for vehicular ad hoc networks.

As we can observe, a general architecture of geographic routing in VANETs is composed of four main layers and two additional modules to treat the geographic information. The architecture could be seen as a top-down approach. The first layer is named *Application Layer*, which is used to offer VANET applications as interfaces between users and communication layers. The *Transport Layer* can operate using traditional transport protocols (i.e. TCP or UDP) as well as specific transport protocol for VANETS (i.e. Car-2-X [13]). In the *Network Layer* we find the services and procedures provided by the geographic routing protocol, such as the location services and the forwarding procedures. The last

[1]Sometimes it is called the back up mode or recovery mode in the literature [7] [9]

 ISBN: 978-1-61208-118-2

layer is the *Physical Layer*, which can be operational using conventional wireless communication protocols (i.e. IEEE 802.11a/b/g) and VANET wireless communication protocols (i.e. 802.11p [14]).

Besides that, additional modules are vertically added in the left and right sides of the architecture, namely, *Information Connector* and *Management*. The *Information Connector* operates as a cross-layer approach to support efficient and structured information exchange across the layers. Likewise, the *Management* module is able to manage this information for improving robustness and reliability of packet delivery in vehicular communication.

In spite of the existence of these layers and modules, our focus in this paper is related to geographic routing protocols in VANETs. Therefore, we continue presenting the services and procedures contained in the *Network Layer*. As previously explained, a geographic routing protocol is generally consisted of a location service and a geographic forwarding strategy, which are described in the Sections IV and V.

IV. Location Service

Kasemann et al. [15] introduced an example of location service, which was called Reactive Location Service (RLS). When the source node executes the RLS, it sends a message to discover the position of the destination, containing the identification of the destination node in addition to its identification and position. The message is flooded in the network until the destination is reached or the Time-To-Live expires. At the moment that the message is received by the destination, a response of localization is sent to the source node, containing the position of destination node [16].

In [17], the authors presented the Vehicle Location Service (VLS) protocol, a map-based vehicle location service for city environments. In summary, the information in digital maps is used to perform location service. They present a new method of partitioning the network and constructing distributed location servers.

Another strategy to put in practice the location service can be provided by cellular and infrastructure networks, where each node notifies its position to a specific server that stores location information for a set of nodes. In the proposal presented in [18], each node initially communicates to each other one of its current location. The server for a node is defined as the set of nodes located within a circle of limited radius, centered at its initial position. Before sending a location update message the node geocasts such a message to its server. In other words, the location update message is unicasted using geographic routing until it reaches one node inside the server. This message is then disseminated inside the server's circle. When the destination node is found, the source node sends out two search messages. The first is oriented to the last known position of the destination, and the second is forwarded in direction of the destination's server. When the search message arrives at the destination or a node within the server, the source is notified with the current location [19]. Once the destination's position is known, the geographic routing protocol initiates the geographic forwarding.

V. Geographic Forwarding

As previously described, geographic forwarding algorithms work in two different modes: greedy mode and void-handling mode. The principal difference leads in determining the situation where it is convenient to use each. The greedy forwarding mode is used whenever possible, and the void-handling mode is strictly used when the greedy forwarding mode cannot be applied.

A. Greedy Forwarding

Several greedy-forwarding algorithms perform different optimization techniques to select the next-hop node closest to the destination node. Face routing [20] is one of the fundamental algorithms for routing packets using compass routing on geometric networks. The key idea is to select the best path along the faces intersected by the line segments between a source and a destination. To avoid loops using face routing, it is required a planar graph of the original network. A planar graph can be defined as a sub graph without crossing edges, which represents the same connectivity as the original network. Therefore, face routing guarantees to reach the destination after as long as the network topology is a planar graph [21].

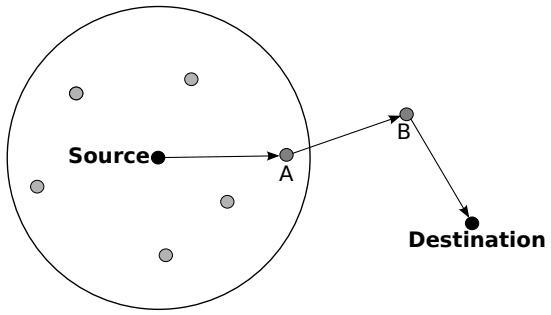


Figure 2. Greedy forwarding mode.

Figure 2 shows an example of the greedy forwarding mode. As we can observe, the *Source* node sends a message to the *Destination* node. Once the location service detects the node positions, the greedy routing algorithm selects the next-hope closest to the destination, for example, the node *A*. Then, the node *A* chooses the next-hope using the same selection rules until the message reach the *Destination* node. When a node cannot locate the next-hope node, it can use the void-handling mode [7]. In this model, the node decides to route the packet around the void since there is a possibility that a valid path from the source to the destination node exists.

 ISBN: 978-1-61208-118-2

B. Void-handling Mode

The void-handling mode is strictly applied as a recover strategy to deliver packets when the greedy forwarding mode cannot be used due the existence of communications voids. As cited in the Section II, a void occurs as a result of high signal interference in the communication due to the presence of large buildings. These buildings or lack of radio coverage can result in voids in the physical network topology. Therefore, voids can obstruct the packet forwarding process at local minima, where the neighbors close to the destination are hidden/unreachable, resulting in a failure.

Communications voids are considered as a serious problem for any feasible geographic routing protocol. It is important to know how to handle voids in an effective and efficient manner. Besides that, it is a difficult task to predict when and where a void will occur due to the unpredictable patterns of node deployment and the uncertain dynamics of time varying wireless network environments. Thus, data packets can be lost in the network if a robust void-handling strategy is not implemented, wasting network resources as well as disabling communications between pair of nodes.

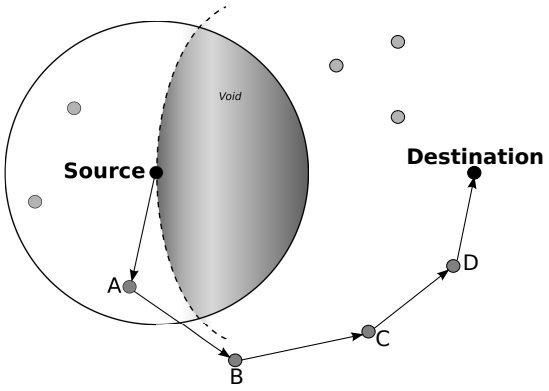
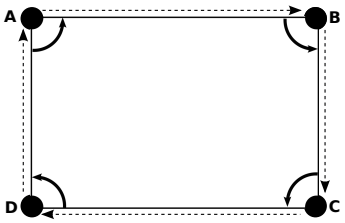


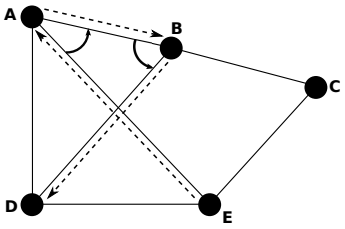
Figure 3. Void-handling mode.

Figure 3 presents an example of void-handling technique due to occurrence of a communication void. According to the example, a *Source* node desires to send a packet to *Destination* node. However, the *Source* node is closer to the *Destination* than any of its neighboring nodes that are located within its radius. Consequently, it is not possible to send a packet using the greedy forwarding mode. In this case, the packet is said to have encountered a communications void and can be sent by the path (*A - B - C - D*). Finally, the *Source* node is called a void node while the shaded region, without any nodes inside, is a void area.

The simplest void-handling strategy is *flooding* the network from the *Source* node to all neighboring nodes. If each node executes the same procedure, this strategy will certainly enable the packet to reach the *Destination* node when at least a path is found. However, this strategy is effective but inefficient in relation to resource utilization, since each node has to forward the packet and the *Destination* node may receive many copies of the same data packet from different paths.



(a) Correct operation.



(b) Fail.

Figure 4. Right Hand Rule.

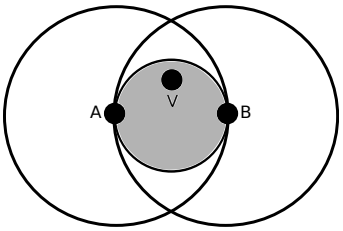
VI. Challenges of using geographic routing in VANETs

There are several prerequisites on the availability of position information in VANET environments, such as position-awareness of each participating vehicle, e.g., a GPS receiver installed on every vehicle. However, this assumption of using position systems is possible due the multiplication of Global Position System (GPS) and the progress on self-configuring localization mechanisms in urban scenarios. Thus, it is important that each vehicle be aware of each neighbor position. A way to perform the position updates is sending beacon messages that indicate the current position of the vehicle.

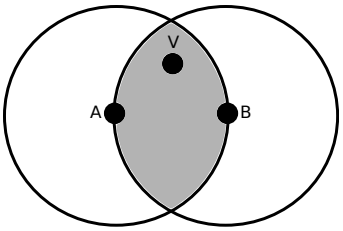
The Greedy Perimeter Stateless Routing (GPSR) protocol is one of the most important algorithms to demonstrate the basic concepts of geographic routing in vehicular ad hoc networks. There are several proposals that use GPSR models to offer new geographic protocols in VANET scenarios, such as [22] and [23]. In summary, the GPSR is a purely local decision strategy, since no route setup or maintenance is required. Instead, forwarding hops are determined 'on the fly' [2]. It applies both greedy forwarding to send packets using position information and a void-handling technique through the perimeter mode as a recover strategy when the greedy forwarding fails. In such a case, position information points in the right direction but is not correlated with available paths to the destination.

Two void-handling techniques are used in the perimeter mode of GPSR protocol. The first technique is similar to an online routing algorithm for planar graphs and the second

 ISBN: 978-1-61208-118-2



(a) GG.



(b) RNG.

Figure 5. Gabriel Graph (GG) and Relative Neighborhood Graph (RNG).

is a distributed planarization algorithm. The algorithm for planar graphs is applied using the Right Hand Rule as presented in Figure 4. As we can observe in Figure 4(a), if a packet starts from the node *A* in direction to the node *B*, the next node will be *C*, following the edges *(A,B)* and *(B,C)*. Thus, the complete in the graph is $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$. The use of planar graphs is a good strategy for recovering the local maxima problem[2]. By definition, the planar graphs take into account those pairs whose edges do not cross or intersect in the plane. However, in the example presented in Figure 4(b), the graph that represents the VANET scenario is not a planar graph. Consequently, the GPSR utilizes a planarization algorithm.

According to Figure 4(b), if the path starts in the edge *(E,A)* and the Right Hand Rule is applied, the path will be composed by $E \rightarrow A \rightarrow B \rightarrow D$. Therefore, the node C is not reached. Two planarization algorithms are utilized in the GPSR: Relative Neighborhood Graph (RNG) and Gabriel Graph (GG) [24]. In summary, they perform removing algorithms to generate connected RNG and GG graphs. Based on Figure 5(a), a GG occurs if an edge (A, B) is found and there are not other vertices with diameter equal to the distance between A and B as well as do not cross these two vertices. On the other hand, if we observe Figure 5(b), a RNG graphs occurs when an edge (A, B) is found and the distance between A and B is minor or equal to the distance among a vertex V and A or B. Hence, we can conclude that a RNG graph is a sub graph of a GG graph.

The use of planarization algorithms in vehicular ad hoc networks allows performing the Right Hand Rule due to the elimination of crossing edges. However, the elimination algorithm can remove essential nodes in the VANET scenario, which can result in the network disconnection. Other problem of using planarization algorithms is related to the excessive number of hops. For example, a vehicle can directly send a packet to the destination, but it sends to the next-hop most close to the destination. Other problems are associated to routing loops and wrong directions. The routing loops occur, for example, when a source node is in the right side of its two-hops neighbor node in the graph. Likewise, the wrong direction problem is found when there is more than one routing alternative, resulting in the use of long routes to deliver the packets.

To avoid planar graphs problems, Lochert et al. created the Geographic Source Routing (GSR) [2]. The key idea is to use the information contained in digital maps to compute routes, which creates an overlay network. The route is calculated using the Dijkstra algorithm. Similarly to GSR, Tian et al. proposed the Spatially Aware packet Routing (SAR) [25]. They use an association among digital maps and graphs. The routing process is based on the source routing.

In [26] the authors presented a geographic routing to avoid routing loop problems in urban VANETs. GeoCross exploits the natural planar feature of urban maps without resorting to cumbersome planarization. Its feature of dynamic loop detection makes GeoCross suitable for highly mobile VANET. The same authors presented in [27] a new approach using delay and disruption tolerant strategies, which is a hybrid geographic routing solution enhancing the standard greedy and recovery modes exploiting the vehicular mobility and on-board vehicular navigation systems to efficiently deliver packets even in partitioned networks.

VII. Conclusion

With the widespread adoption of Global Position System (GPS) and the progress on self-configuring localization mechanisms, geographic routing in VANET has garnered significant attention to provide promising solutions for message delivery. The main motivation to study routing protocols in VANETs is related to the expansion of data exchange among vehicles in order to provide robust applications for Intelligent Transportation Systems (ITS). VANET applications can include on-board active safety systems, providing communications among nearby vehicles (V2V) and between vehicles and the roadside infrastructure (V2I). However, several challenges have to be overcome before application on a large scale.

This paper presented main geographic routing strategies in VANETs scenarios. It shows benefits of use location-aware routing and discusses strengths and weaknesses of those approaches. Finally, the issues that need further investigations in this area were discussed.

References

[1] Y.-B. Wang, T.-Y. Wu, W.-T. Lee, and C.-H. Ke, "A novel geographic routing strategy over vanet," apr. 2010, pp. 873 –879.

[2]This problem occurs when the *Source* node is closer to the *Destination* than any of its neighboring nodes that are located within its radius.

 ISBN: 978-1-61208-118-2

[2] C. Lochert, H. Hartenstein, J. Tian, H. Fussler, D. Hermann, and M. Mauve, "A routing strategy for vehicular ad hoc networks in city environments," jun. 2003, pp. 156 – 161.

[3] I. Stojmenovic, "Position-based routing in ad hoc networks," *Communications Magazine, IEEE*, vol. 40, no. 7, pp. 128 – 134, jul. 2002.

[4] M. Mauve, A. Widmer, and H. Hartenstein, "A survey on position-based routing in mobile ad hoc networks," *Network, IEEE*, vol. 15, no. 6, pp. 30 –39, nov. 2001.

[5] Z. Jin, Y. Jian-Ping, Z. Si-Wang, L. Ya-Ping, and L. Guang, "A survey on position-based routing algorithms in wireless sensor networks," *Algorithms*, vol. 2, no. 1, pp. 158–182, 2009. [Online]. Available: http://www.mdpi.com/1999-4893/2/1/158/

[6] H. Cheng and J. Cao, "A design framework and taxonomy for hybrid routing protocols in mobile ad hoc networks," *Communications Surveys Tutorials, IEEE*, vol. 10, no. 3, pp. 62 –73, 2008.

[7] D. Chen and P. Varshney, "A survey of void handling techniques for geographic routing in wireless networks," *Communications Surveys Tutorials, IEEE*, vol. 9, no. 1, pp. 50 –67, 2007.

[8] B. Karp and H. T. Kung, "Gpsr: greedy perimeter stateless routing for wireless networks," in *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*. New York, NY, USA: ACM, 2000, pp. 243–254.

[9] M. Heissenbuttel, T. Braun, T. Bernoulli, and M. Wachli, "Blr: beacon-less routing algorithm for mobile ad hoc networks," *Computer Communications*, vol. 27, no. 11, pp. 1076–1086, 2004.

[10] M. Zorzi and R. Rao, "Geographic random forwarding (geraf) for ad hoc and sensor networks: energy and latency performance," *Mobile Computing, IEEE Transactions on*, vol. 2, no. 4, pp. 349 – 365, oct. 2003.

[11] C. Lemmon, S. M. Lui, and I. Lee, "Geographic forwarding and routing for ad-hoc wireless network: A survey," in *NCM '09: Proceedings of the 2009 Fifth International Joint Conference on INC, IMS and IDC*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 188–195.

[12] Q. Yang, A. Lim, S. Li, J. Fang, and P. Agrawal, "Acar: Adaptive connectivity aware routing protocol for vehicular ad hoc networks," aug. 2008, pp. 1 –6.

[13] K. David and A. Flach, "Car-2-x and pedestrian safety," *Vehicular Technology Magazine, IEEE*, vol. 5, no. 1, pp. 70 –76, mar. 2010.

[14] S.-Y. Wang and C.-C. Lin, "Nctuns 5.0: A network simulator for ieee 802.11(p) and 1609 wireless vehicular network researches," sep. 2008, pp. 1 –2.

[15] W. Kiess, H. Fussler, J. Widmer, and M. Mauve, "Hierarchical location service for mobile ad-hoc networks," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 8, no. 4, pp. 47–58, 2004.

[16] T. Camp, J. Boleng, and L. Wilcox, "Location information services in mobile ad hoc networks," vol. 5, 2002, pp. 3318 – 3324 vol.5.

[17] X.-y. Bai, X.-m. Ye, J. Li, and H. Jiang, "Vls: a map-based vehicle location service for city environments," in *ICC'09: Proceedings of the 2009 IEEE international conference on Communications*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 2694–2698.

[18] J. Widmer, M. Mauve, H. Hartenstein, and H. Fubler, "Position-based routing in ad hoc wireless networks," pp. 219–232, 2003.

[19] M. C. Weigle, S. Olariu, M. Abuelela, and G. Yan, "Use of Infrastructure in VANETs," in *Vehicular Networks: From Theory to Practice*, S. Olariu and M. C. Weigle, Eds. USA: Chapman & Hall/CRC, 2009.

[20] E. Kranakis, S. O. C. Science, H. Singh, and J. Urrutia, "Compass routing on geometric networks," in *in Proc. 11 th Canadian Conference on Computational Geometry*, 1999, pp. 51–54.

[21] J. You, Q. Han, D. Lieckfeldt, J. Salzmann, and D. Timmermann, "Virtual position based geographic routing for wireless sensor networks," *Comput. Commun.*, vol. 33, no. 11, pp. 1255–1265, 2010.

[22] S. Rao, M. Pai, M. Boussedjra, and J. Mouzna, "Gpsr-l: Greedy perimeter stateless routing with lifetime for vanets," oct. 2008, pp. 299 –304.

[23] S. Sun, J. Kim, Y. Jung, and K. Kim, "Zone-based greedy perimeter stateless routing for vanet," in *ICOIN'09: Proceedings of the 23rd international conference on Information Networking*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 374–376.

[24] T.-H. Su and R.-C. Chang, "Computing the constrained relative neighborhood graphs and constrained gabriel graphs in euclidean plane," *Pattern Recogn.*, vol. 24, no. 3, pp. 221–230, 1991.

[25] J. Tian, L. Han, and K. Rothermel, "Spatially aware packet routing for mobile ad hoc inter-vehicle radio networks," vol. 2, oct. 2003, pp. 1546 – 1551 vol.2.

[26] K. C. Lee, P.-C. Cheng, and M. Gerla, "Geocross: A geographic routing protocol in the presence of loops in urban scenarios," *Ad Hoc Netw.*, vol. 8, no. 5, pp. 474–488, 2010.

[27] P.-C. Cheng, K. C. Lee, M. Gerla, and J. Harri, "Geodtn+nav: Geographic dtn routing with navigator prediction for urban vehicular environments," *Mob. Netw. Appl.*, vol. 15, no. 1, pp. 61–82, 2010.

 ISBN: 978-1-61208-118-2

Envelope Interfaces for Geoscientific Processing with High Performance Computing and Information Systems

Claus-Peter Rückemann
Leibniz Universität Hannover (LUH),
Westfälische Wilhelms-Universität Münster (WWU),
North-German Supercomputing Alliance (HLRN), Germany
Email: ruckema@uni-muenster.de

***Abstract*—The results presented in this paper describe the achievements with the development of techniques, implementing a standard way for computing communication in complex integrated information and computing systems. Compute Envelopes can provide means for generic data processing and flexible information exchange, namely computation objects. Targeting mission critical environments, the interfaces can embed instruction information, validation and verification methods. The application covers challenges of collaborative implementation, legal, and security issues with these processes. A major task is integrating information systems with Distributed and High Performance Computing (HPC) resources in natural sciences disciplines, scientific information systems, for building integrated public / commercial information system components within the e-Society. The main focus of this paper is on how to simplify the integration of information and computing resources from modular system architectures, using envelope techniques reducing complexity and strengthening trust aspects in future integrated information and computing systems.**

Keywords–Computing and Information Systems; Geoscientific Processing; Distributed and High Performance Computing.

1. Introduction

There are two main objectives for interfacing modular complex integrated information and computing systems: “trust in computing” and “trust in information”. This paper concentrates on implementing methods for flexible use of envelope interfaces for use with integrated information and computing systems for managing objects and strengthen trust with systems from natural and geosciences, spatial sciences, and remote sensing as to be used for application, e.g., in environment management, healthcare or archaeology. Spatial means are tools, for the sciences involved. Therefore processing and computing is referred to the content which is embedded and used from the visual domain.

Over the last years a long-term project, Geo Exploration and Information (GEXI) [1] for analysing national and international case studies and creating as well as testing various implementation scenarios, has shown the two trust groups of systems, reflected by the collaboration matrices [2]. It has examined chances to overcome the deficits, built a collaboration framework and illuminated legal aspects and benefits [3]. The information and instructions handled within these systems is one of the crucial points while systems are evolving by information-driven transformation [4].

For computing and information intensive systems the limiting constraints are manyfold. Recycling of architecture native and application centric algorithms is very welcome. In order to reuse information about these tasks and jobs, it is necessary to enable users to separate the respective information for system and application components. This can be done by structured envelope-like descriptions containing essential workflow information, algorithms, instructions, data, and meta data. The container concept developed has been called Compute Envelope (CEN). The idea of envelope-like descriptive containers has been inspired by the good experiences with the concept of Self Contained applications (SFC) [5]. Envelopes can be used to integrate descriptive and generic processing information. Main questions regarding the topics of computing envelope interfaces are: Which content can be embedded or referenced in envelopes? How will these envelope objects be integrated into an information and computing system and how can the content be used? How can the context and environment be handled?

This paper is organised as follows. Section 2 presents motivation and implementation challenges. Section 3 describes the implementation complexity. Sections 4 and 5 show the basics of envelope files, architecture, and meta data aspects. Sections 6 and 7 describe the processes of compute access and report on the implementation for integrated systems. Sections 8, 9, and 10 present the evaluation and summarise the lessons learned, conclusion, and outlook.

2. Motivation

The complexity of integrated information and computing systems is an interesting factor that makes it well worth applying meta-instructions and signatures for algorithms and interfaces. Within these systems users will have some defined characteristics, so that these can be described in a flexible way. The economic profit is even greater in case of application with informations systems handling huge numbers of objects. In these cases, envelopes can be used to provide a standard interface for computing access.

The purpose of CEN files is to get recyclable algorithms

 ISBN: 978-1-61208-118-2

and information for defined system environments. If the algorithms handled for computing purposes would concentrate on ordinary processing batch scripts only, this would be pointless and tedious and anyone would need his own computing instructions. Todays information and computing systems are facing challenges from complex environments and heterogeneous content. These present aggravation conditions for any long term implementation.

3. Implementation complexity and amazements

There is a number of scenarios described in the previous publications [1], [2], [3], [6], [7], showing how "trust in computing" and "trust in information" can more easily be achieved by reducing complexity for development partners in otherwise very complex systems. Examples of elements for data objects being subject to handling and protection are:

- Vector data and multi-dimensional data,
- Raster data (aerial, remote sensing, and photographic),
- Primary and secondary spatial information,
- Calculation, measurement, and processing results,
- Meta data, instruction and interactive information,
- Commercially provided or licensed data, etc.

Most problems arise from complexity necessary to reflect the use cases and being built on prepackaged components each having own practical 'amazements' for integrated development and from content and context handling.

3.1. Problems for complex use cases

At todays state of development, available technology is still very limited if we take a look on long term issues, stability, and extendability. Integrated information and computing systems components make use of various technologies: batch systems, schedulers, IPC, sandboxing, embedded applications, browser plugins, remote execution, network and communication protocols, computing interfaces as well as public and sensitive data. The major motivation is to create an architecture of system components to ease the use of various content in different resources environmental context.

3.2. Content and context

The most important experience found from the case studies over the last years is, that besides the algorithmic content can be signed, the context of the CEN cannot be signed in any way. So, due to the complexity of the environment, the fitness for a specific purpose cannot be guaranteed.

What can be controlled is given in the following list. For some cases these characteristics can even be signed.

- Information and computing resources instructions,
- Links between the information and computing system,
- Prerequisites of the computing system,
- Processing directives and script elements,
- Input / output data necessary, etc.

What cannot be fully validated and 'signed' is environment and network specifications, nodes characteristics, state of the components of the system, ...

The signer of a script would have to create and sign a compute envelope for a defined system context. This is comparable to signature for document files. These only refer to the content namely the object and not for the context outside, where the data will be used, namely the environment and action. The parts of the algorithm that cannot be signed, can in many cases be checked for logical and consistent state. The environment has to be validated and verified with the installation and configuration of the information and computing system. Looking at these aspects it would be beyond the scope of this paper as this will be future work.

4. Compute instructions into envelopes

SFC applications are used for over ten years now, but there has been no general concept for communication in integrated environments. These applications can hold data, executable binary and bytecode, components, configuration, libraries and other parts in a way supporting flexible and portable use. Even more flexibility can be achieved for non binary components and executables, meaning non system dependent elements. Starting from the proof of concept, a more suitable solution has now been created on a generic envelope base. An end-user public client application may be implemented via a browser plugin, based on appropriate services. The current solution is based on CEN files containing XML structures for handling and embedding compute information. Common envelopes that have been used with content signing, are Object Envelope (OEN) files. Envelope files can even be integrated with sources like GISIG Active Source, Active Map (GAS, GAM), with Object Oriented (OO) support, embedded into SFC applications. For various applications the use from within binary executables or TBC has been considered for any application part and data [5]. The case studies have shown good results as porting components can be made easier and application service providers and users can react faster on system modifications. Listing 1 shows a generic CEN file.

```
1  <ComputeEnvelope><!-- ComputeEnvelope (CEN)-->
2  <Instruction>
3  <Filename>Processing_Bat_GIS515.torque</Filename>
4  <Md5sum>...</Md5sum><Sha512sum>...</Sha512sum>
5  <DateCreated>2010-08-01:231523</DateCreated>
6  <DateModified>2010-08-01:232734</DateModified>
7  <ID>...</ID><CertificateID>...</CertificateID>
8  <Signature>...</Signature><Content>...</Content>
9  </Instruction>
10 </ComputeEnvelope>
```

Listing 1. Example for a Compute Envelope (CEN).

Listing 2 shows a small example using DataReference instead of embedded data inside a CEN.

```
1  ...<Content><DataReference>https://doi...</
    DataReference></Content>...
```

Listing 2. CEN using DataReference.

 ISBN: 978-1-61208-118-2

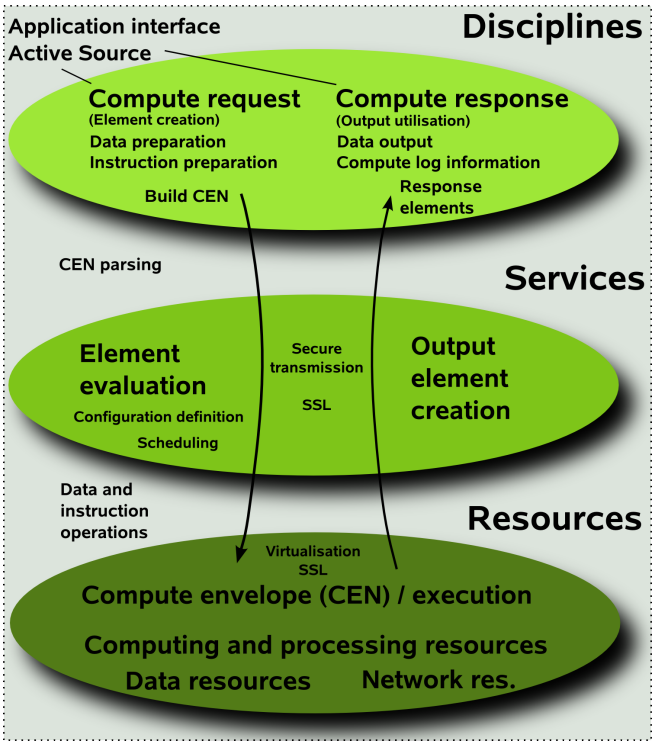


Figure 1. Compute access for integrated systems.

Content can be handled as content-stream or as content-reference (Listing 2). The way this will have to be implemented for different use cases depends on the situation, an in many cases on the size and number of data objects. One benefit of content-reference with high performant distributed or multicore resources is that references can be processed in parallel on these architectures. The number of physical parallel resources and the transfer capacities inside the network are limiting factors.

5. Fundamental implementation architecture

The fundamental architecture is based on a layered concept for the implementation and operation of information and computing systems [3]. Currently "trust in computing" can cover the content aspects. Context aspects are out of scope with todays systems. For the three development layers this mainly states tasks for services and resources layers.

Figure 1 shows the compute access for integrated information and computing systems. Providing application support for integrated systems does base on two assumptions, that applications can be modified to support an integration of compute resources and that the configuration of the system resources infrastructure can be configured in a suitable way. This means extending the application sources on one hand and to enable automating of access and workflow cycles. The meta structure can be used to store the instruction sets and to describe the job and the environment. Various meta data is necessary to describe the signed object data. For chronology as well as for plausibility, the security of the time and data association is important. Integrated system components as well as interested parties must be able to use this meta data as well as means must be available to verify that the time stamps associated with an object are authentic and hold integrity (trusted time stamp authorities). This may be a function of the CA, respective a dedicated time server service. The envelope is able to describe any form of embedded or referred meta data.

6. Compute access for integrated systems

From the experiences of the case studies, the following passages describe the workflow (Figure 1) necessary to implement integrated information and computing systems in context with envelopes interfaces.

Based on the three layers of the Grid-GIS house framework, namely disciplines, services, and resources, the compute access is established by several steps, consisting of a compute request block and a compute response block.

6.1. The request process

Starting with the application interface a compute request starts with the element creation. The main parts consist of preparation of the data needed for computing or processing and the preparation of the necessary instructions. In the case studies Active Source applications have been used. From these prerequisites the CEN are built.

- **Disciplines layer:** The user's compute request is started based on computation data and the instruction information necessary for processing. Elements for the CEN are created by the user. The CEN is built from the elements, supported by application functions.
- **Services layer:** The elements are evaluated and adapted by the system configuration definition. The instruction sets are prepared for scheduling.
- **Resources layer:** Data and instruction operations are handled by batch system or interactive use. Compute, data, network, and storage resources are used with elements and configuration by services layer definition.

Depending on the system configuration, architecture, and allowed user behavior, the processing can run interactively or in batch mode. This can take from seconds to days for elements, depending on the purpose of the integrated system concept, before data will be suitable to be delivered for the response process.

6.2. The response process

With the request process delivering output, the following steps take place:

- **Resources layer:** The resulting output will be handled as described in the CEN instructions. Very large data can be stored on appropriate storage resources for later use, smaller or interactive data can be directly delivered to the services layer.

 ISBN: 978-1-61208-118-2

- **Services layer:** Services functions handle the output and do create output elements, delivered to the user or interface defined in the original CEN envelope.
- **Disciplines layer:** The data from the output elements is delivered for utilisation to the user or interface, e.g., to be interactively integrated into the application.

7. Implemented solution for integrated systems

For most interactive information system components a configuration of the distributed resources environment was needed. In opposite to OEN use, making it necessary to have referenced instead of embedded data for huge data sets, for CEN it should be possible to embed the essential instruction data in most cases. So there is less need for minimising data overhead and communication. Envelope technology is meant to be a generic extensible concept for information and computing system components. Figure 2 shows the workflow with application scenarios from the GEXI case studies. Future objectives for client components are:

- Channels for limiting communication traffic,
- Qualified signature services and accounting,
- Using signed objects without verification,
- Verify signed objects on demand.

The tests done for proof of concept have been in development stage. A more suitable solution has now been created on a generic envelope base. An end-user public client application may be implemented via a browser plugin, based on appropriate services. The current solution is based on CEN files containing XML structures for handling and embedding data and information.

7.1. Integrated components in practice

When taking a look onto different batch and scheduling environments one can see large differences in capabilities, handling different environments and architectures. In the last years experiences have been gained in handling simple features for different environments for High Throughput Computing like Condor [8], workload schedulers like LoadLeveler [9] and Grid Engine [10], and batch system environments like Moab / Torque [11], [12]. Batch and interactive features are integrated with Active Source event management [5]. Listing 3 shows a small example of a CEN embedded into an Active Source component.

```
#BCMT-------------------------------------------------
###EN \gisigsnip{Object Data: Country Mexico}
#ECMT-------------------------------------------------
proc create_country_mexico {} {
global w
#  Sonora
$w create polygon 0.938583i 0.354331i 2.055118i ...
#BCMT-------------------------------------------------
###EN \gisigsnip{Compute Data: Compute Envelope (CEN)}
#ECMT-------------------------------------------------
#BCEN <ComputeEnvelope>
##CEN <Instruction>
##CEN <Filename>Processing_Bat_GIS515.torque</Filename>
##CEN <Md5sum>...</Md5sum>
##CEN <Sha1sum>...</Sha1sum>
##CEN <Sha512sum>...</Sha512sum>
##CEN <DateCreated>2010-09-12:230012</DateCreated>
##CEN <DateModified>2010-09-12:235052</DateModified>
##CEN <ID>...</ID><CertificateID>...</CertificateID>
##CEN <Signature>...</Signature>
##CEN <Content>...</Content>
##CEN </Instruction>
#ECEN </ComputeEnvelope>
...
proc create_country_mexico_autoevents {} {
global w
$w bind legend_infopoint <Any-Enter> {set killatleave [
 exec ./mexico_legend_infopoint_viewall.sh $op_parallel
  ] }
$w bind legend_infopoint <Any-Leave> {exec ./
 mexico_legend_infopoint_kaxv.sh }
$w bind tulum <Any-Enter> {set killatleave [exec
 $appl_image_viewer -geometry +800+400 ./
 mexico_site_name_tulum_temple.jpg $op_parallel ] }
$w bind tulum <Any-Leave> {exec kill -9 $killatleave }
} ...
```

Listing 3. CEN embedded with Active Source.

Interactive applications based on Active Source have been used on Grid, Cluster, and HPC (MPP, SMP) systems [6].

7.2. Resources interface

Using CEN features, it is possible to implement resources access on base of validation, verification, and execution. The sources (Listing 4, 5) can be generated semi-automatically and called from a set of files or can be embedded into an actmap component, depending on the field of application.

```
<ComputeEnvelope><!-- ComputeEnvelope (CEN)-->
<Instruction>
<Filename>Processing_Batch_GIS612.pbs</Filename>
<Md5sum>...</Md5sum>
<Sha1sum>...</Sha1sum>
<Sha512sum>...</Sha512sum>
<DateCreated>2010-08-01:201057</DateCreated>
<DateModified>2010-08-01:211804</DateModified>
<ID>...</ID>
<CertificateID>...</CertificateID>
<Signature>...</Signature>
<Content><DataReference>https://doi...</DataReference><
 /Content>
<Script><Pbs>
<Shell>#!/bin/bash</Shell>
<JobName>#PBS -N myjob</JobName>
<Oe>#PBS -j oe</Oe>
<Walltime>#PBS -l walltime=00:10:00</Walltime>
<NodesPpn>#PBS -l nodes=8:ppn=4</NodesPpn>
<Feature>#PBS -l feature=ice</Feature>
<Partition>#PBS -l partition=hannover</Partition>
<Accesspolicy>#PBS -l naccesspolicy=singlejob</
 Accesspolicy>
<Module>module load mpt</Module>
<Cd>cd $PBS_O_WORKDIR</Cd>
<Np>np=$(cat $PBS_NODEFILE | wc -l)</Np>
<Exec>mpiexec_mpt -np $np ./dyna.out 2>&1</Exec>
</Pbs></Script>
</Instruction>
</ComputeEnvelope>
```

Listing 4. Embedded Active Source MPI script.

Examples for using High Performance Computing and Grid Computing resources include batch system interfaces and job handling. Job scripts from this type will on demand (event binding) be sent to the batch system for processing. The Actmap Computing Resources Interface (CRI) is an example for an actmap library (`actlcri`) containing functions and

 ISBN: 978-1-61208-118-2

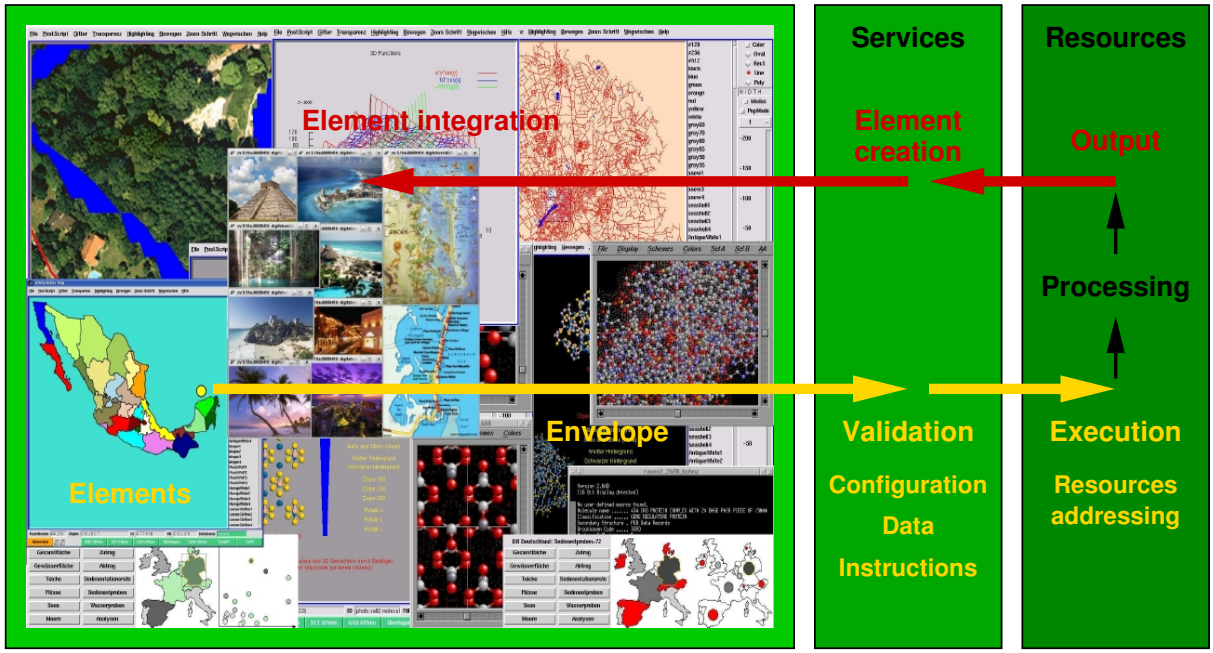


Figure 2. Workflow with application scenarios from the GEXI case studies.

procedures and even platform specific parts in a portable way. CRI can be used for handling computing resources, loading Tcl or TBC dynamically into the stack [13] when given `set behaviour_loadlib_actlib "yes"`.

```
<ComputeEnvelope><!-- ComputeEnvelope (CEN)-->
<Instruction>
<Filename>Processing_Batch_GIS612.pbs</Filename>
<Md5sum>...</Md5sum>
<Sha1sum>...</Sha1sum>
<Sha512sum>...</Sha512sum>
<DateCreated>2010-08-01:201057</DateCreated>
<DateModified>2010-08-01:211804</DateModified>
<ID>...</ID>
<CertificateID>...</CertificateID>
<Signature>...</Signature>
<Content><DataReference>https://doi...</DataReference><
 /Content>
<Script><Condor>
<Environment>universe = standard</Environment>
<Exec>executable = /home/cpr/grid/job.exe</Exec>
<TransferFiles>should_transfer_files = YES</
 TransferFiles>
<TransferInputFiles>transfer_input_files = job.exe,job.
 input</TransferInputFiles>
<Input>input = job.input</Input>
<Output>output = job.output</Output>
<Error>error = job.error</Error>
<Log>log = job.log</Log>
<NotifyMail>notify_user = ruckema@uni-muenster.de</
 NotifyMail>
<Requirements>
requirements = (Memory >= 50)
requirements = ( ( (OpSys=="Linux")||(OpSys=="AIX"))&&(
 Memory >= 500) )
</Requirements>
<Action>queue</Action>
</Condor></Script>
</Instruction>
</ComputeEnvelope>
```

Listing 5. Embedded Active Source Condor script.

With Actmap CRI being part of Active Source, calls to parallel processing interfaces, e.g., using InfiniBand, can be used, for example MPI (Message Passing Interface) and OpenMP, already described for standalone job scripts for this purpose, working analogical. [7].

8. Evaluation

Targets for handling computing content were integrated information and computing systems. It has been possible to describe and handle components from prepackaged software, interfaces, and architectures within the envelopes for generic processing of geo- and related data. The primary benefits of the solution using CEN as stated from the case studies are:

- Build a defined interface between dedicated information system components and computing system components.
- Unique algorithm for using environment components.
- Integration of information and computing systems.
- Speed-up the development of new and modification of existing components in complex environments.
- Portable, transparent, extendable, flexible, and scalable.
- Hierarchical structured meta data, easily parsable.
- OO-support (object, element) on application level.
- Multi-system support.
- Support for validation and verification of object elements via Public Key Infrastructure (PKI).
- Usable with sources and binaries like Active Source.

Main drawbacks are:

- Additional layer.
- Complexity of parsing and configuration.

The context is an important aspect, though it cannot be called “drawback” here. With closed products, e.g., when memory requirements are not transparent, it is difficult for users to specify their needs. Anyhow, testing is in many cases not

 ISBN: 978-1-61208-118-2

the answer in productive environments. Separate measures have to be taken to otherwise minimise possible problems and ease the use of resources in productive operation.

Even in the face of the drawbacks, for information systems making standardised use of large numbers of accesses via the means of interfaces, the envelopes can provide efficient management and access, as programming interfaces can.

9. Lessons learned

CEN based on generic envelopes has provided a very flexible and extensible solution for creating portable, secure processing components with integrated information and computing systems. The case study showed that nearly any instruction set and data structure can be handled with CEN in embedded or referred use. Instruction information, meta data, signatures, and check sums can be used and customised in various ways for implementing information and computing system components. Support for transfer and staging of data in many aspects further depends on system configuration and resources as for example physical bottlenecks cannot be eliminated by any kind of software means. For future integrated information and computing systems an interface layer between user configuration and system configuration would be very helpful. From system side in the future we need working and least operation-invasive operating system resources limits, e.g., for memory and a flexible limits management.

10. Conclusion and future work

The envelopes have been successfully used for various Active Source components, for all data objects in use, embedded and referenced. Objects can be integrated in various native ways so the content is available as 'standard' content. As the context of the system environment existing, cannot be comprehensively defined, this leads to some interesting consequences. Exchange between systems is quite expendable without common standards. What we would need for future systems is some virtualised processing sandbox for integrating batch and scheduling systems, providing standardised definable interfaces. This concept has demonstrated a basic approach in order to begin to pave the way for information and computing systems and show the next aspects to go on with.

Acknowledgements

We are grateful to all national and international academic and industry partners in the GEXI cooperations for the innovative constructive work and to the colleagues at the Leibniz Universität Hannover, IRI, HLRN, WWU, ZIV, D-Grid for participating in fruitful case study examples and the participants of the EULISP Programme for prolific discussion of scientific, legal, and technical aspects over the last years.

References

[1] "Geo Exploration and Information (GEXI)," 1996, 1999, 2010, URL: http://www.user.uni-hannover.de/cpr/x/rprojs/en/index.html#GEXI (Information) [accessed: 2010-05-02].

[2] C.-P. Rückemann, "Legal Base for High End Computing and Information System Collaboration and Security," *International Journal on Advances in Security*, vol. 3, no. 3&4, 2010, (to appear), ISSN: 1942-2636, URL: http://www.iariajournals.org/security/ [acc.: 2010-10-11].

[3] C.-P. Rückemann, "Legal Issues Regarding Distributed and High Performance Computing in Geosciences and Exploration," in *Proceedings of the Int. Conf. on Digital Society (ICDS 2010), The Int. Conf. on Technical and Legal Aspects of the e-Society (CYBERLAWS 2010), February 10–16, 2010, St. Maarten, Netherlands Antilles*. IEEE Computer Society Press, IEEE Xplore Digital Library, 2010, pp. 339–344, ISBN: 978-0-7695-3953-9, URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5432414 [acc.: 2010-03-28].

[4] M. Mackert, P. Whitten, and B. Holtz, *Health Infonomics: Intelligent Applications of Information Technology*. IGI Global, 2009, pp. 217–232, chapter XII, in: Pankowska, M. (ed.), Infonomics for Distributed Business and Decision-Making Environments: Creating Information System Ecology, ISBN: 1-60566-890-7, DOI: 10.4018/978-1-60566-890-1.ch010.

[5] C.-P. Rückemann, "Beitrag zur Realisierung portabler Komponenten für Geoinformationssysteme. Ein Konzept zur ereignisgesteuerten und dynamischen Visualisierung und Aufbereitung geowissenschaftlicher Daten," Diss., Westfälische Wilhelms-Universität, Münster, Deutschland, 2001, 161 (xxii+139) S., URL: http://wwwmath.uni-muenster.de/cs/u/ruckema/x/dis/download/dis3acro.pdf [accessed: 2009-11-16].

[6] C.-P. Rückemann, "Geographic Grid-Computing and HPC empowering Dynamical Visualisation for Geoscientific Information Systems," in *Proceedings of the 4th International Conference on Grid Service Engineering and Management (GSEM), September 25–26, 2007, Leipzig, Deutschland, co-located with Software, Agents and services for Business, Research, and E-sciences (SABRE 2007)*, R. Kowalczyk, Ed., vol. 117. GI-Edition, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik e.V. (GI), 2007, pp. 66–80, ISBN: 78-3-8579-211-6, ISSN: 1617-5468.

[7] C.-P. Rückemann, "Dynamical Parallel Applications on Distributed and HPC Systems," *International Journal on Advances in Software*, vol. 2, no. 2, 2009, ISSN: 1942-2628, URL: http://www.iariajournals.org/software/ [accessed: 2009-11-16].

[8] "Condor, High Throughput Computing," 2010, URL: http://www.cs.wisc.edu/condor/ [accessed: 2010-10-10].

[9] "IBM Tivoli Workload Scheduler LoadLeveler," 2005, URL: http://www-03.ibm.com/systems/software/loadleveler/ [accessed: 2010-10-10].

[10] "Sun Grid Engine," 2010, URL: http://gridengine.sunsource.net/ [accessed: 2010-10-10].

[11] "Moab: Admin Manual, Users Guide," 2010, URL: http://www.clusterresources.com [acc.: 2010-10-10].

[12] "Torque Admin Manual," 2010, URL: http://www.clusterresources.com/torquedocs21/ [accessed: 2010-10-10].

[13] "Tcl Developer Site," 2010, URL: http://dev.scriptics.com/ [accessed: 2010-10-10].

 ISBN: 978-1-61208-118-2

An Integrated Geospatial Data Management System in a Complex Public Research Environment using Free and Open Source Software

Christian Braun, Ulrich Leopold
Public Research Centre Henri Tudor,
Resource Centre for Environmental Technologies,
Esch-sur-Alzette, Grand Duchy of Luxembourg
e-mail: Christian.Braun@tudor.lu, Ulrich.Leopold@tudor.lu

***Abstract*—The interdisciplinary nature of environmental research centres, dealing with geospatial data, analysis and environmental modelling on a daily basis, requires specific methods and technologies in the field of geospatial information management. The large amount of generated information has to be stored, catalogued, visualised and treated effectively for further analysis. The Public Research Centre Henri Tudor has set up a prototype system to create an integrated geospatial data infrastructure, serving the needs of various user profiles from novice level to advanced and experienced data analysts and modellers. The paper will show solutions on how to give a broad range of users access to an integrated infrastructure. This is achieved by introducing different user interfaces: an easy to use web interface for beginners - advanced web mapping and feature services coupled to desktop GIS applications for intermediates - direct data base access, making use of cutting-edge geospatial tools and spatially distributed modelling algorithms for experts. The system is fully functional on all user levels and based on free and open source software. It is integrating current standards of the Open Geospatial Consortium, to assure exchange with stakeholders and to guarantee its further functional extensibility.**

Keywords - Geospatial data management; geospatial data infrastructure; Web services; Web GIS; INSPIRE.

I. INTRODUCTION

Since the enactment of the Infrastructure for Spatial Information in Europe (INSPIRE) directive [1] in 2007 the development of geospatial information services has increased. Stakeholders are obliged to create functional data and software infrastructures and to be compliant to defined data exchange guidelines. Web-enabled geospatial technologies are key components [2]; data base management systems with components to store spatial objects [3], data access libraries [4], map rendering engines [5] and web front ends are constantly adapted and developed to meet different needs.

These components, though interactive, are merely data viewers; the information flow has to be more open and bidirectional when users operate with the data by further geospatial analysis or by creating new data sets. In this case the focus is on the interaction between data storage, single software components and modelling tools. If a geospatial data management infrastructure has to handle operations like real-time mapping or automated interpolations [6], then further components have to be integrated. Examples for these are algorithms powered by geostatistical software tools like R and Gstat [7] or GIS software libraries that are working on data base level like TerraLib [8].

In the recent years, the Public Research Centre Henri Tudor has faced an increasing amount of geospatial information, generated and compiled by its daily work for more and more demanding models. This called for integrated geospatial data management techniques and a concept to design a solution based on free and open source software. The solution to this is a geospatial data infrastructure that is able to operate as a central hub, serving geospatial information to provide the necessary data to each collaborator.

This paper reflects on the possibilities of this proposition to build a fully functional integrated geospatial data infrastructure with free and open source components by the given constraints and requirements. This will be described in the next section, followed by an extensive description of used components and how they interact in Section III. Finally, a short description is given in Section IV on how the system is used in the institute so far and which major improvements will have to be made in the future.

II. SYSTEM DESIGN, REQUIREMENTS AND CONSTRAINTS

The topic of geospatial data was quite new at the institute and the collaborators were introduced to existing techniques and demo applications on the Internet illustrating how workflows could look like with advanced data management tools. In several brainstorming sessions certain requirements were identified and a prototype was designed. This led to a fully functional geospatial data infrastructure with advanced data management capabilities.

One major constraint was that the whole infrastructure had to be built on free and open source software products. The main aim of this is to reduce licensing costs and to increase interoperability and extensibility considering open

 ISBN: 978-1-61208-118-2

standards. With the interdisciplinary nature of an environmental research institute the system should be fully scalable to serve the needs of the various user profiles in particular.

An easy to use web based spatial information system should be an entry point for exploration (meta data search) and visualisation of all connected data pools. Furthermore, expert users should be able to make use of advanced geospatial data analysis and modelling with connected programming environments such as R [9], shown in Figure 1, and geographic information systems like GRASS GIS [10]; both embedded in high performance computing facilities.

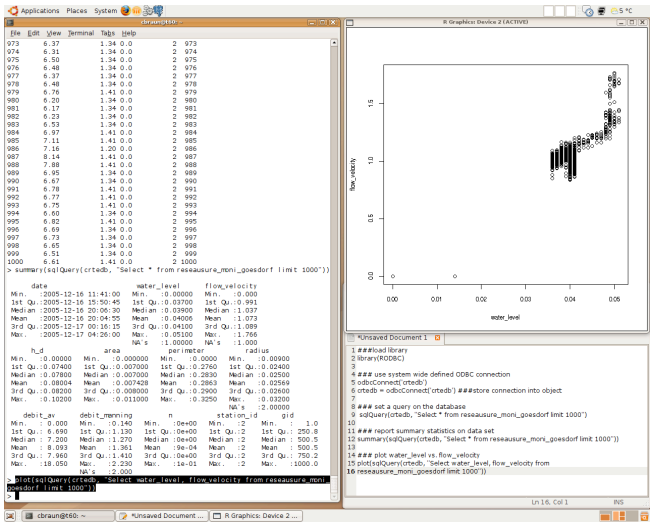


Figure 1. R implemented as modelling tool.

Another component is the meta data management and data exchange in standardised formats as defined by the Open Geospatial Consortium (OGC) [11]. This ensures the compliance with infrastructures on the national level (e.g. GeoPortal Luxembourg [12]) and the fulfilment of international reporting needs compatible with the INSPIRE directives. Meta data are a crucial part to clearly identify data sets and have information about extent, quality, spatial and temporal schema, spatial reference, and distribution of digital geographic data. One wide spread standard and "best practice" is ISO 19115 [13] with its XML implementation schema ISO 19139 [14]. In addition, the general effort in maintenance and supervision of the whole system should be as small as possible.

These requirements and constraints were leading to a first sketch of the infrastructure design shown in Figure 2 summarising the most important parts of the geospatial data infrastructure.

The central part is represented by the physical data storage system, which consists of one or more data bases and provides interfaces for exchange with software clients, such as GIS and other modelling environments used in the research centre. An application server is coupled to the physical storage and represents the high performance computing facilities of the institute. For data exploration, administration and maintenance the geospatial data infrastructure should provide a meta data system as well as administration tools.

An example workflow is the geospatial analysis and automated mapping of field measured soil properties (e.g. texture, content of organic carbon and water uptake rate) using automated mapping algorithms as developed in the INTAMAP project [15]. A user can upload the relevant data sets via a web interface to the central data base where he is able to set corresponding access privileges to his working group. If the gathered data does not need further analysis in a laboratory, field devices with compatible software could upload it via cellular network without effort. Later, different group members are able to access the data with the connected software products of their choice for an initial overview mapping or to check data for consistency. After positive data checks, further geostatistical analysis, such as interpolation with kriging methods [16] will be done and the data as well as analysis results are fed into the meta data base and the web mapping interface. This ensures fast access of necessary up-to-date information to all collaborators.

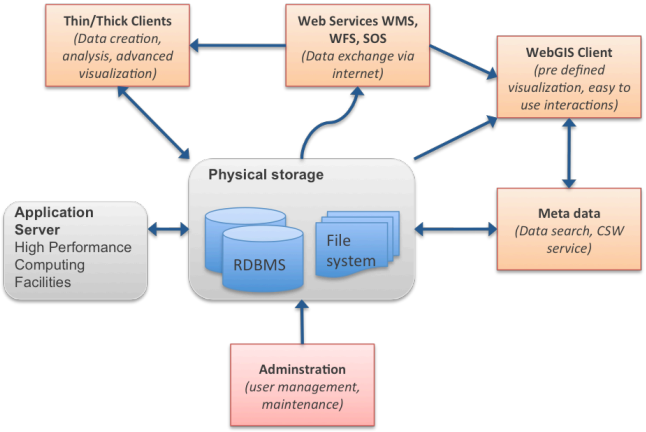


Figure 2. Overview of the main components of the prototype design.

With the implementation of Web Processing services (WPS) [17] users are able to access and run predefined models on the application server. Necessary data sets and parameters have to be included in the server request. This is a favourably tool for non-advanced users to run complex calculations with automated error checking capabilities and visualise outputs directly in the WebGIS as overlay maps. This is a recent achievement running a cast shadow calculation module in a city area implemented in PyWPS [18]. We are using the OGC WPS standard which offers native support for GRASS GIS and supports a generic WPS Java Script library.

III. THE GEOSPATIAL DATA INFRASTRUCTURE

The design of the geospatial data infrastructure can be seen in Figure 3, with special consideration of the given constraints and user needs. The figure shows the used software components and their arrangement and connection from a technical point of view. It is organized in five different layers with their unique assigned functions in the system.

 ISBN: 978-1-61208-118-2

The data pools at the bottom are representing the physical storage of different data sets. Vector and meta data are stored in a PostgreSQL [19] data base system with PostGIS [20] extension to allow the storage of spatial objects as well. Currently, raster data is stored simply file based, with ideas to move and couple it to the above mentioned PostgreSQL set up driven by PostGIS Raster [21]. This project is about to implement the raster data type as much as possible like vector type data in PostGIS. It will also offer a single set of overlay SQL functions operating seamlessly on vector and raster data. General user access and data security is handled on data base level by user login and password protection. With user and group privileges on table level inside the data base a granular access right system can be established to protect sensitive information.

Libraries like GDAL/OGR [22] make sure that access and coordinate transformation of geospatial objects is possible. A map server like MapServer [23] is able to create different OGC compliant services, such as Web Map Service (WMS), Web Feature Service (WFS) and Web Coverage Service (WCS) [18] and is providing the rendering of maps for the web client p.mapper [24]. This web interface is mainly based on PHP MapScript and is accessing data sets directly via the MapServer rendering engine.

Direct data access for data analysis software, such as the wide spread Microsoft Excel, or the more advanced environments like R and S-Plus could be realised through ODBC/JDBC interfaces. Desktop GIS clients, like GRASS GIS, uDig [25] or Quantum GIS [26] can access the data storage directly with built-in interfaces to visualize or edit geospatial data sets.

Another important component of the infrastructure is a catalogue service for meta data and a corresponding web interface like GeoNetwork [27]. It allows the user to interactively search for sets of any information in the system via related meta data. Furthermore it is a crucial part of compatibility to other infrastructures because GeoNetwork is acting as a CSW (Catalogue Service Web) service to publish data sets in higher-level infrastructures (e.g. Geoportal Luxembourg, INSPIRE services). Further it can connect to them as client as well. This feature enables harvesting of foreign data sets to link them into the own system by OGC web services.

The geospatial data infrastructure and its diverse options of data access enable the acquisition and compilation of relevant information in an integrated way for all types of user levels. It ensures easy and compatible data exchange and it is fully functional and well integrated in the daily work flow of collaborators.

A main tool for data research is the WebGIS and the web based meta data application where most basic information on available data can be retrieved. A most fundamentally work flow, such as search, selection, visualisation of geospatial information and printing of basic maps can easily be done via the web map interface adding full support to basic or novice users. Users with advanced geocomputation skills are using the WebGIS itself mainly for discussion purposes during meetings. Nevertheless, they are creating and analysing data sets by making use of powerful server hardware and further installed software products like PostGIS, which is offering spatial data queries and geoprocessing methodologies for vector data directly implemented into SQL (Structured Query Language).

Further applications could be automated mapping of environmental parameters or statistical analysis tools built into the web based map interface to provide further insight to data to the more inexperienced user. The Public Research Centre Henri Tudor is also involved in geospatial uncertainty analysis and modelling within the developments of the UncertWeb project [28] and its Uncertainty Markup Language (UncertML). These are of importance in the near future, e.g. in decision making at administrative levels.

All functional blocks are built on free and open source software (FOSS) components. FOSS is used in many applications and operational infrastructures, making use of advantages like the constant and continuous development and on-going support by the developer community. The monetary benefit of free and open source software could be reduced in the beginning phase because of extra expenses caused by an additional need in development of features that are not available out-of-the-box. In terms of desktop GIS software, standardized interfaces to OGC services are available due to the effectiveness of OGC standards and the active community support.

IV. CONCLUSION AND FUTURE WORK

The presented infrastructure has the following benefits: improved data organisation and management, high data security and fast availability of information on all user levels.

The system is able to connect all GIS and modelling tools that are used in the institute seamlessly and present them in an integrated way to the user.

As European frameworks like INSPIRE are just being implemented, the integration of distributed data sources will continue in the future. This integration will be gradual, adding data when it is available and prepared for integration. The system provides all necessary interfaces to achieve an easy and straightforward integration of all OGC compliant data sources. At the same time it also complies with all OGC standards to provide data for many other geospatial data infrastructures and applications.

Future work will have to be done in the fields of a more seamless WPS integration and the set up of more calculation modules to assist collaborators. The storage of raster data should be integrated in the data base system by the recent development of PostGIS Raster. This would allow making use of basic raster map algebra, also in combination with vector data, without using dedicated GIS tools. Furthermore, the overall usability of the web frontend should be improved with better support for data queries and quick look summary statistics of available data sets. In addition, the system should be referenced in the national geospatial data infrastructure to assure an exchange of information and make use of meta data services.

 ISBN: 978-1-61208-118-2

REFERENCES

[1] Directive 2007/2/EC of the European Parliament and of the council of 4 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), http://inspire.jrc.ec.europa.eu 05/10/2010

[2] Anderson, G. and Moreno-Sanchez, R. (2003): Building web-based spatial information solutions around open specifications and open software, Transactions in GIS, vol. 7(4), pp. 447-466.

[3] Yeung, A.K.W. and Brent Hall, G. (2007): Spatial database systems: design, implementation and project management. Springer.

[4] Warmerdam, F. (2008): The Geospatial Data Abstraction Library. In: Open Source Approaches to Spatial Data Handling, vol. 2, pp. 87-104. Springer, Berlin.

[5] Vatsavai, R., Shekhar, S., Burk, T., and Lime, S. (2006): UMN-MapServer: A High-Performance, Interoperable, and Open Source Web Mapping and Geo-spatial Analysis System. In: Geographic, Information Science. Lecture Notes in Computer Science, vol. 4197, pp. 400-417. Springer, Berlin.

[6] Brenning, A. and Dubois, G. (2007): Towards generic real-time mapping algorithms for environmental monitoring and emergency detection. Stochastic Environmental Research and Risk Assessment, vol. 22(5), pp. 601-611.

[7] Bivand,R.S., Pebesma, J.E., and Gómez-Rubio, V. (2008): Applied spatial data analysis with R. Springer.

[8] Câmara, G., Vinhas, L., Ferreira, K., Queiroz, G., Souza, R.C.M., Monteiro, A.M. et al. (2008): TerraLib: An open-source GIS library for large-scale environmental and socio-economic applications. In: Open Source Approaches to Spatial Data Handling, vol. 2, pp. 247-270. Springer, Berlin.

[9] R Development Core Team, 2010, R: A Language and Environment for Statistical Computing, Vienna, Austria. http://www.R-project.org 05/10/2010

[10] GRASS Development Team, 2010, Geographic resources analysis support system (GRASS GIS) software, Trento, Italy. http://grass.osgeo.org 05/10/2010

[11] Open Geospatial Consortium, http://opengeospatial.org 05/10/2010

[12] Administration du Cadastre et de la Topographie, http://act.public.lu 01/10/2010

[13] ISO 19115:2003, http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020 06/12/2010

[14] ISO 19139:2007, http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557 06/12/2010

[15] Interoperability and Automated Mapping: INTAMAP, http://www.intamap.org 29/09/2010

[16] Chilès, J.-P. and Delfiner, P. (1999): Geostatistics: modeling spatial uncertainty. Wiley, New York.

[17] OGC Standards and Services, http://www.opengeospatial.org/standards 10/10/2010

[18] PyWPS, http://pywps.wald.intevation.org/ 06/12/2010

[19] PostgreSQL, http://www.postgresql.org 10/10/2010

[20] PostGIS, http://postgis.refractions.net 10/10/2010

[21] PostGIS Raster, http://trac.osgeo.org/postgis/wiki/WKTRaster 10/10/2010

[22] GDAL, http://www.gdal.org 15/09/2010

[23] MapServer, http://mapserver.org 10/10/2010

[24] p.mapper, http://pmapper.net 15/09/2010

[25] uDig – User-friendly Desktop Internet GIS, http://udig.refractions.net 03/09/2010

[26] Quantum GIS, http://www.qgis.org 15/09/2010

[27] GeoNetwork opensource, http://geonetwork-opensource.org 05/10/2010

[28] UncertWeb, http://www.uncertweb.org 12/10/2010

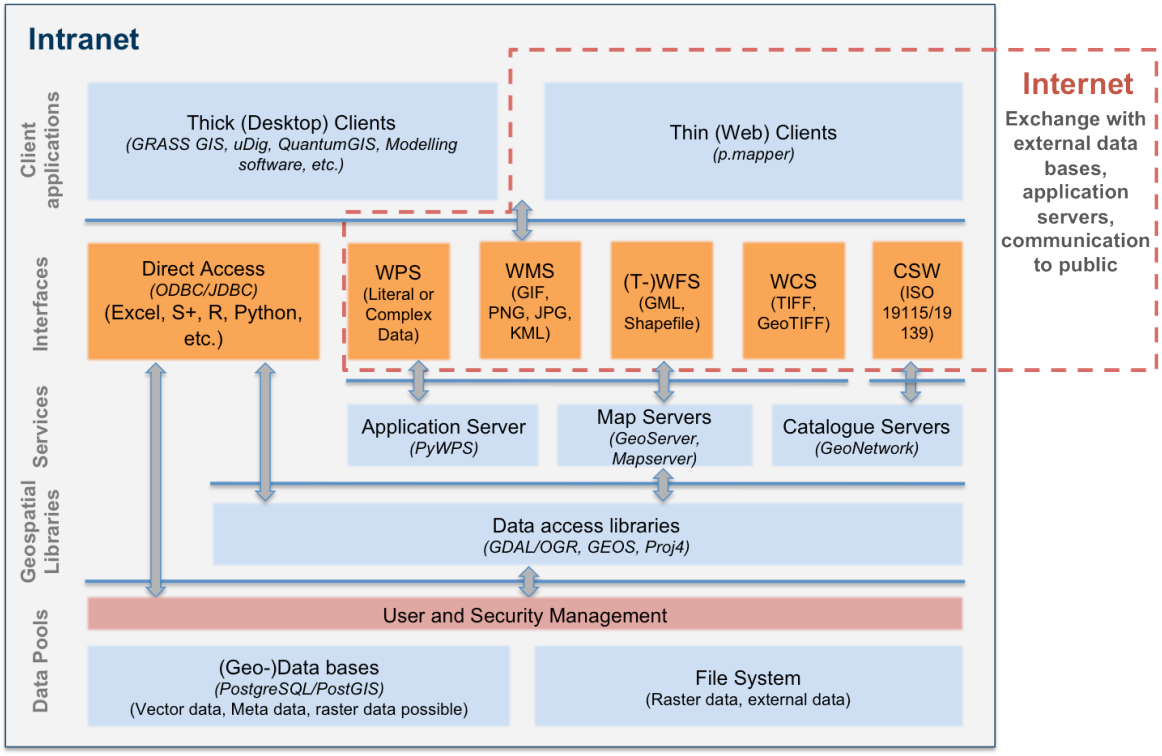


Figure 3. Detailed schema of components of the geospatial data infrastructure.

Analysis of Radio Communication Attenuation Using Geoprocessing Techniques

Mehdi Mekni
Department of Computer Science
Sherbrooke University
Sherbrooke, Canada
mmekni@gmail.com

Bernard Moulin
Department of Computer Science and Software Engineering
Laval University
Quebec, Canada
bernard.moulin@ift.ulaval.ca

***Abstract*—Multi-Agent Geo-Simulation (MAGS) aims to simulate phenomena involving a large number of autonomous situated actors (implemented as software agents) evolving and interacting within a Virtual representation of the Geographic Environment (VGE). A radio communication system is a typical complex dynamic phenomena where transmitter and receiver antennas are constantly constrained by the physical environment in which it they are deployed. In the real world, radio transmissions are subject to propagation effects which deeply affect the received signals because of geographic and environmental characteristics (foliage and vegetation, buildings, mountains and hills, etc.). Using geoprocessing techniques, we propose an automated approach to build semantically-informed and geometrically-accurate virtual geographic environments which uses Geographic Information System (GIS) data and builds an informed graph-based structure called Informed Virtual Geographic Environment (IVGE). In addition, we propose a multi-agent prototype to analyze the attenuation effect due to the radio signal's traversal between antennas (simulated as software agents) through terrain shape, vegetation area, and buildings using a 3D line-of-sight computation technique.**

***Keywords-Informed Virtual Geographic Environment (IVGE)*; *Radio Signal Propagation*; *Line-Of-Sight*; *Multi-Agent Geo-Simulation (MAGS)*.**

I. INTRODUCTION

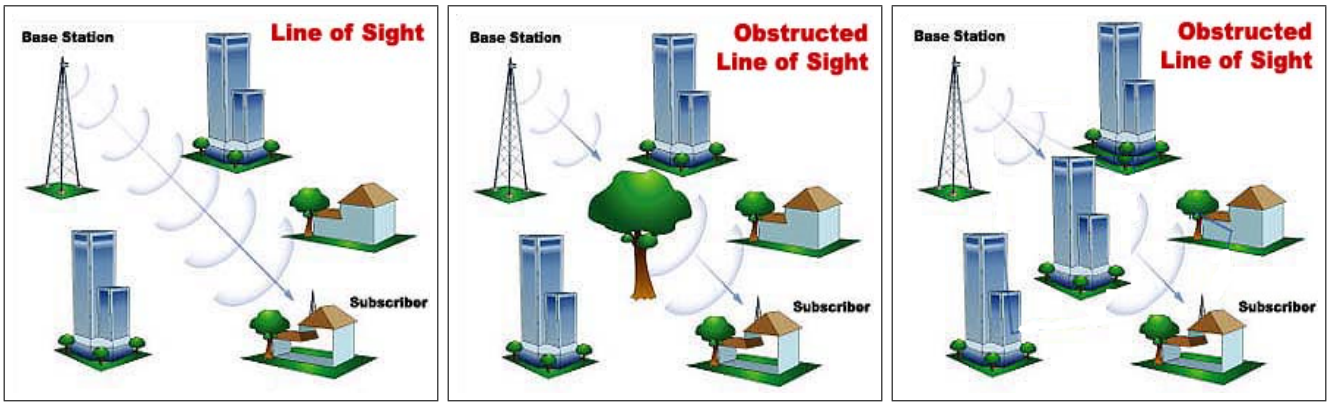
During the last decade, the Multi-Agent Geo-Simulation (MAGS) approach has attracted a growing interest from researchers and practitioners to simulate phenomena in a variety of domains including traffic simulation, crowd simulation, urban dynamics, and changes of land use and cover, to name a few [2]. Such approaches are used to study phenomena (i.e., car traffic, mobile robots, mobile networks, crowd behaviours, etc.) involving a large number of simulated actors (implemented as software agents) of various kinds evolving in, and interacting with, an explicit description of the geographic environment called Virtual Geographic Environment (VGE). Nevertheless, simulating such autonomous situated agents remains a particularly difficult issue, because it involves several different research domains: geographic environment modeling, spatial cognition and reasoning, situation-based behaviours, etc. When examining situated agents in a VGE, whether for gaming or simulations purposes, one of the first questions that must be answered is how to represent the world in which agents navigate [20]. Since a geographic environment may be complex and large-scale, the creation of a VGE is difficult and needs large quantities of geometrical data describing the environment characteristics (terrain elevation, location of objects and agents, etc.) [16] as well as semantic information that qualifies space (buildings, roads, parks, etc.) [17]. Current approaches usually consider the environment as a monolithic structure, which considerably limits the way that large-scale, real world geographic environments and agent's spatial reasoning capabilities are handled [13].

Rapid advances in wireless communications have made mobile data applications a high-growth area of development. So far, most applications only focus on geographic data collection and access using geographic information systems. However, many emergency management applications need such geographic data in order to ensure that field workers and command center operators collaborate under acceptable operating conditions [7]. Under emergency conditions, emergency systems need to quickly establish an *ad hoc*, ground-level network of radio stations mounted on temporary command centers, vehicles, or temporary masts, interacting with moving field operators using mobile devices [6].

Radio Frequency (RF) communication does have some limitations that must be considered. The maximum line-of-sight range between two shoulder-height devices is limited to $12km$ considering the curvature of the earth but not considering refraction of radio waves. The actual range may be considerably less depending on transmission power and receiver sensitivity, and the radio signal can be attenuated or degraded due to obstruction resulting from its interactions with features on its transmission path. In an urban environment, possible obstructions include buildings, trees, and bridges for example (*Figure 1*).

The difficulty of evaluating an ad hoc radio network with hundreds of nodes and various levels of mobility operating in a complex geographic environment (i.e., rugged terrain, dense foliage, buildings, etc.) motivated us to use our IVGE model to analyze communication attenuation. In order to

 ISBN: 978-1-61208-118-2



(a) (b) (c)

Figure 1: Radio signal propagation: (a) an obstruction-free propagation; (b) propagation obstructed by vegetation and foliage; and (c) propagation obstructed by buildings.

analyze the communication attenuation in real geographic environments, we propose to virtually reproduce the actual geographic environment using an automated IVGE generation approach [12] and leverage its enriched description to precisely compute the radio transmission's attenuation. The radio transmission is computed using the line of sight between two points located in the IVGE.

In this project, our goal is to analyze the radio communication attenuation in complex, dynamic and large-scale geographic environments. In order to achieve such a goal, a geographic environment model should precisely represent geographic features. It should also integrate several semantic notions characterising these geographic features. Since we deal with large-scale geographic environments, it would be appreciable to have a VGE organized efficiently in order to reduce the spatial computation algorithms such as line-of-sight algorithms. There is also a need for autonomous situated agents representing antennas either transmitters or receivers which are able to communicate using radio signal propagation in presence of both *static* and *dynamic* obstacles located in the VGE. Static obstacles correspond to areas that affect radio signal propagation such as walls, fences, trees, rivers, etc. Static obstacles also include obstructions resulting from terrain elevation (mountains, hills, etc.). Dynamic obstacles correspond to moving or stationary antennas which may interfere with the radio communication system.

In this paper, we propose a novel approach to model and simulate radio communication networks using multi-agent systems evolving in and interacting with a virtual geographic environments. The rest of the paper is organized as follows: in Section II, we provide a brief overview on related work in the field of environment representation and analysis of radio communications. Section III presents our methodology for the creation of informed virtual geographic environments. Section IV highlights the unique properties of our IVGE model which easily and efficiently enable spatial reasoning algorithms and geometrical computations such as line-of-sight computation. Section V details the radio communication analysis tool. Finally, Section VI concludes and presents our perspectives.

II. Related Works

In this section, we provide a brief overview of prior works related to *environment representation*, and *analysis of radio communications* in virtual environments.

A. Environment Representation

Virtual environments and spatial representations have been used in several application domains. For example, Thalmann *et al.* proposed a virtual scene for virtual humans representing a part of a city for graphic animation purposes [5]. Donikian *et al.* proposed a modelling system which is able to produce a multi-level data-base of virtual urban environments devoted to driving simulations [9]. Ali *et al.* used a multi-agent geo-simulation approach to simulate customers' behaviours in shopping malls [1]. More recently, Shao *et al.* proposed a virtual environment representing the New York City's Pennsylvania Train Station populated by autonomous virtual pedestrians in order to simulate the movement of people [16]. However, since the focus of these approaches is computer animation and virtual reality, the virtual environment usually plays the role of a simple background scene in which agents mainly deal with geometric characteristics. Indeed, the description of the virtual environment is often limited to the geometric level, though it should also contain topological and semantic information for other types of applications. Therefore, most interactions between agents and the environment are usually simple, only permitting to plan a path in a 2D or 3D world with respect to free space and obstacle regions [4].

 ISBN: 978-1-61208-118-2

B. Analysis of Radio Communications

Signal specialists currently have limited capabilities for predicting radio performance in complex geographic environments that may involve jammers and may have combinations of open terrain and obstructed locations [6]. Foliated areas, buildings, or terrain may cause obstructions to the radio line-of-sight path [7]. Therefore, there is a need for a communication analysis tool that helps fill this void. We propose to use geoprocessing techniques in order to build a tool for the analysis of radio communications attenuation involving a geometrically-accurate and semantically-informed virtual geographic environment model. This tool is an easy to use, stand-alone, GUI-based application that runs on a PC and provides a rich set of functionalities to aid the user to compute the total path loss of a given operational scenario that is directly coupled to an operational area using reliable GIS data.

This tool enables the user to plan radio deployments and determine link connectivity using actual radio parameters, taking into account the presence of obstacles, while accounting for the excess attenuation due to terrain, foliage (vegetation), and building obstructions. To compute the total path loss, our tool uses the following parameters: height of transmitter antenna (meters), height of receive antenna (meters), transmitter antenna position (x, y, z), receiver antenna position (x, y, z), and frequency of operation (GHz).

III. Computation of IVGE Data

In this section, we present our automated approach to compute the IVGE data directly from vector GIS data. This approach is based on four stages which are detailed in this section (*Figure 2*): *input data selection*, *spatial decomposition*, *maps unification*, and finally, the generation of the *informed topologic graph*.

GIS Input Data Selection: The first step of our approach consists of selecting the different vector data sets which are used to build the IVGE. The only restriction concerning these data sets is that they must respect the same scale. The input data can be organized into two categories. First, *elevation layers* contain geographical marks indicating absolute terrain elevations. As we consider 2.5*D* IVGE, a given coordinate cannot have two different elevations, making it impossible to represent tunnels for example. Multiple elevation layers can be specified, and if this limitation is respected, the model can merge them automatically. Second, *semantic layers* are used to qualify various types of data in space. Each layer indicates the physical or virtual limits of a given set of features with identical semantics in the geographic environment, such as roads or buildings. The limits can overlap between two layers, and our model is able to merge the information.

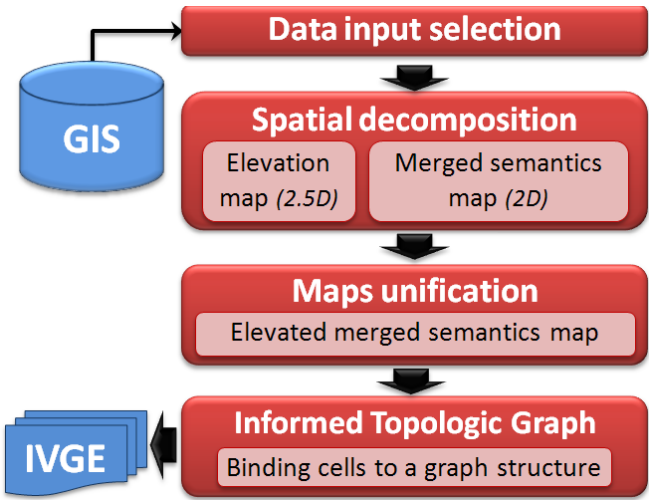


Figure 2: The four stages to obtain an IVGE from GIS data. All the stages are automatic but the first.

Spatial Decomposition: The second step consists of obtaining an exact spatial decomposition of the input data into cells. This process is entirely automatic using Delaunay triangulation, and can be divided into two parts in relation to the previous phase. First, an elevation map is computed, corresponding to the triangulation of the elevation layers. All the elevation points of the layers are injected into a 2D triangulation, the elevation being considered as an attribute of each node. This process produces an environment subdivision composed of connected triangles. Such a subdivision provides information about coplanar areas: the elevation of any point inside a triangle can be deduced by using the elevation of the three original data points to form a plane. Second, a merged semantics map is computed, corresponding to a constrained triangulation of the semantic layers. Indeed, each segment of a semantic layer is injected as a constraint which keeps track of the original semantic data by using an additional attribute for each semantic layer. The obtained map is then a constrained triangulation merging all input semantics. Each constraint represents as many semantics as the number of input layers containing it.

Merging Elevation and Semantics Layers: The third step to obtain our IVGE consists of unifying the two maps previously obtained. This phase can be depicted as mapping the 2*D* merged semantic map onto the 2.5*D* elevation map in order to obtain the final 2.5*D* elevated merged semantics map. First, preprocessing is carried out on the merged semantics map in order to preserve the elevation precision inside the unified map. Indeed, all the points of the elevation map are injected into the merged semantics triangulation, creating new triangles. This first process can be dropped if the elevation precision is not important. Then, a second process elevates the merged semantics map. The elevation of each merged semantics point *P* is computed by retrieving

 ISBN: 978-1-61208-118-2

the corresponding triangle T inside the elevation map, i.e., the triangle whose $2D$ projection contains the coordinates of P. Once T is obtained, the elevation is simply computed by projecting P on the plane defined by T using the Z axis. When P is outside the convex hull of the elevation map then no triangle can be found and the elevation cannot be directly deduced. In this case, we use the average height of the points of the convex hull which are visible from P.

Informed Topologic Graph: The resulting unified map now contains all the semantic information of the input layers, along with the elevation information. This map can be used as an *Informed Topologic Graph* (ITG), where each node corresponds to the map's triangles, and each arc corresponds to the adjacency relations between these triangles. Then, common graph algorithms can be applied to this topological graph, and graph traversal algorithms in particular. One of these algorithms retrieves the node, and therefore the triangle, corresponding to given $2D$ coordinates. Once this node is obtained, it is possible to extract the data corresponding to the position, such as the elevation from the $2.5D$ triangle, and the semantics from its additional attributes. Several other algorithms can be applied, such as path planning or graph abstraction, but they are out of the scope of this paper and will not be detailed here.

IV. Properties of Informed VGE

The subdivision of space into convex cells allows us to preserve the original geometric definition of the geographic environment, unlike the grid-based representations that discretise the environment. Furthermore, the proposed data reorganization produces triangles that feature good properties: convexity which facilitates the geometric calculations; support of heterogeneous geometric constraints (points, segments, polygons); Since each constraint is linked to its nearest neighbor, it is easy to compute the widths of the bottlenecks in the virtual geographic environments. The width computation corresponds to the minimum of borders' width that are not qualified as obstacle.

The subdivision of space into convex cells also allows us to extract an informed topologic graph of the environment featuring relatively few nodes compared to grid-type representations. Additionally, the triangulation is not dependent on a fixed spatial scale for the environment, but only on its complexity (number of constrained segments). It should also be emphasized that curved geometries will produce a lot of triangles since they are represented by a large number of constrained segments. However, since the produced triangulation is represented as a graph, it is possible to abstract it in order to reduce the number of elements. All these properties are of interest to address the issue of line-of-sight computation.

A. Line of Sight's Computation

The spatial subdivision provides a structure of convex cells which facilitates and accelerates the calculation of ray tracing in three dimensions. We define the radius α using the following information: the position of the origin p; the direction vector $\vec{d}$; and the maximum distance considered. Let $Get_{free}(Cell)$ and $Get_{constrained}(Cell)$ be two functions returning respectively the list of free (S_{free}) and constrained (S_{const}) borders bounding the convex cell $Cell$. Let $N(Cell,b)$ be a function returning the normal vector to the border b which belongs to the cell $Cell$ and directed towards the inside of the cell. Finally let us note $\wp(\beta)$ the $2X2$ rotation matrix of $\vec{d}$. The test checking if there is an intersection between the ray and the border b, which links the vertices I and J, is performed using the following expression [11]:

$$\Phi \wedge [\theta \cdot (I-p)) \times \theta \cdot (J-p)))] \leq 0 \quad (1)$$

Where Φ is detailed in equation 2 and θ in equation 3

$$\Phi = \vec{d} \cdot N(Cell,b) \leq 0 \quad (2)$$

$$\theta = \left(\frac{I+J}{2} - p\right) \times \wp\left(\frac{\pi}{2}\right) \quad (3)$$

The line of sight computation algorithm proceeds as follows:

- ***Step 1***: the cell $Cell$ containing the source of the line of sight vector $\vec{LoS}$ is determined.
- ***Step 2***: an intersection test is performed between $\vec{LoS}$ and each border b of $Cell$.
- ***Step 3***: compute $S_{free}(Cell)$ using $Get_{free}(Cell)$ and $S_{const}(Cell)$ using $Get_{constrained}(Cell)$.
- ***Step 4***: if no intersection is found with borders from $S_{free}(Cell)$, then $\vec{LoS}$ must intersect with a border from $S_{const}(Cell)$.
- ***Step 5***: the border b is pushed back to the list of borders crossed by $\vec{LoS}$.
- ***Step 6***: the cell $Cell$ is pushed back to the list of cells crossed by $\vec{LoS}$.
- ***Step 7***: the cell sharing the border b which intersects with $\vec{LoS}$ becomes the current cell. Proceed to Step 2.

The line of sight algorithm, due to its low computational cost, can be extensively used in MAGS involving a large number of agents evolving in a complex IVGE.

V. Analysis of Radio Communication

Planning communications links requires the ability to assess the performance of each link in the presence of a num-

ber of degrading factors. In addition to the normal free space signal attenuation loss, other losses reduce the signal level. These additional losses can be caused by obstructions such as buildings and vegetation (foliage). Analyzing obstruction losses can be difficult because of the geometric, topologic, and semantic characteristics of the geographic environment and the need for path loss models.

We use our ray-tracing algorithm to determine the path that a radio signal takes to arrive at the receiver's position from a given transmitter within our 3D IVGE. We model transmitters and receivers as infinitesimally small points such that paths are computed precisely and cannot be duplicated or missed as sometimes could happen in the approximate approach [15]. Obviously, our ray-tracing algorithm is more precise and reliable. Our approach computes the total source-to-destination path length and then, determines whether the vector defined by the source and destination points (locations in the IVGE) passes through an obstruction area. Doing so, it is able to compute the total path loss between transmitter and receiver antennas. One of three cases may occur: 1) *Obstruction-Free Path*: Vector does not penetrate any obstruction; 2) *Obstruction Block Penetration*: Vector penetrates one or more foliage obstruction blocks and/or one or more building obstruction blocks; and 3) *Ground Penetration*: Vector penetrates the ground (earth) once or several times.

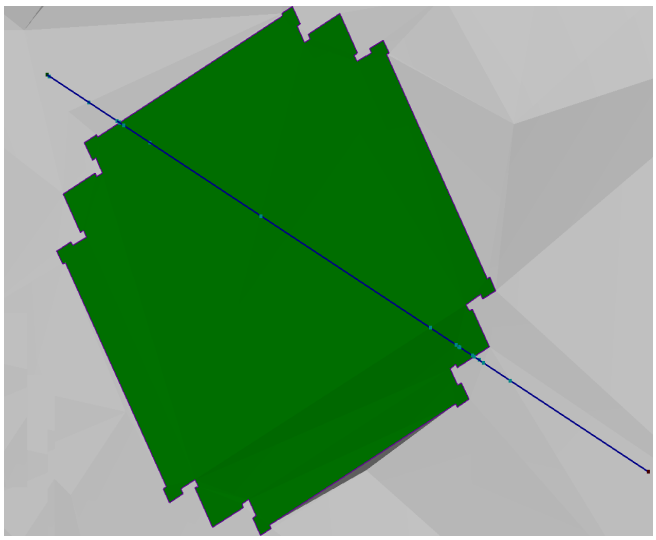


Figure 3: Computation of the ray tracing: the radio signal propagation path (blue), free space (grey), building (green), and vertices representing the intersection locations (light blue).

The computations performed by our tool to quantify the path attenuation for each of the three cases defined above are based on the following mathematical models described in [3].

Plane-Earth Attenuation Model: Let L_p be the path attenuation using the plane-Earth model (dB):

$$L_p = 40Log(D) - 20Log(H_t) - 20Log(H_r) \quad (4)$$

where D is the total source-to-destination path length (meters), and H_t and H_r are the heights of the transmitter and receiver antennae above ground level, respectively (meters).

Free-Space Attenuation Model: Let L_{fs} be the path attenuation using the Free-Space model (dB):

$$L_{fs} = 32.45 + 20Log(D) + 20Log(f) \quad (5)$$

where D is the total source-to-destination path length (meters) and f is the RF frequency (GHz).

Obstruction Block Penetration Model: Let L_B be the path attenuation term due to propagation through a building obstruction block (dB) [14]:

$$L_B = K_1(0.6)^f + K_2 D_B \quad (6)$$

where f is the RF frequency (GHz), K_1 is a constant used to map the first expression above to building penetration data reported in [3]. $K_1 = 35$; K_2 is a constant to account for the attenuation (per meter) of the signal within the building. $K_2 = 1$ (dB/m); and D_B is the distance that the signal propagates through the building (meters).

The first term in Equation 6 accounts for the penetration into and out of the building by the signal and was derived from data reported in [3] using a regression analysis technique. The second term in the equation above accounts for the attenuation through the building and is based on data reported by Willassen in [19].

Foliage obstruction Attenuation Model: If the penetration is through a foliage obstruction block, the tool computes an excess path attenuation term called L_f using the Wiessberger model [18] as follows:

$$L_f = 1.33 f^{0.284}.D_f^{0.588}, 14 < D_f \quad (7)$$

$$L_f = 0.45 f^{0.284}.D_f^{1.0}, \quad 0 \leq D_f \leq 14m \quad (8)$$

where D_f is the distance that the signal propagates through the foliage obstruction (meters) and f is the RF frequency (GHz).

The term 'excess attenuation' refers to the additional attenuation above the basic transmission loss, for a given path length, in the absence of foliage. Our analysis approach applies equations 4 to 8 to calculate the basic transmission loss for the link. Thus, the total path attenuation for the obstruction penetration case called L_{total} is calculated as follows:

$$L_{total} = max(L_p, L_{fs})_{Deff} + sum(L_f) + sum(L_B) \quad (9)$$

 ISBN: 978-1-61208-118-2

Figure 4: Simulation of radio communications' attenuation; yellow lines correspond to obstacle-free line-of-sight radio signal propagation; red lines correspond to obstructed line-of-sight; (1) represents the transmitter antenna implemented using the agent paradigm; (2) an example of a plane-earth obstruction; (3) an example of a block-penetration obstruction.

where $max(L_p, Lfs)_{Deff}$ means that this term is calculated at $Deff$; $Deff$ is the effective path length over which the *Plane-Earth* and *Free-Space* path attenuation model is applied and is equal to the total path length minus the sum of the building obstruction block path length segments or sum(DB); $sum(L_f)$ is the sum of the excess attenuation terms in dB due to signal propagation through the foliage obstruction block(s) (dB); and $sum(L_B)$ is the sum of the path attenuation terms due to propagation through the building obstruction block(s) (dB). *Figure 4* illustrates the agent-based simulation tool that we developed in order to implement our approach to analyze the radio signal attenuation in informed virtual geographic environments. This figure presents a snapshot of the simulation at time t_0 with an agent representing a transmitter antenna and several agents representing receiver antennae.

VI. Discussion and Conclusion

Using reliable GIS data along with the line of sight algorithm (ray tracing feature) provided by our IVGE allows the system to compute the exact locations of intersections occurring between the radio signal propagation path and the terrain shape (*Figure 4*). Our IVGE also allows us to collect the list of cells crossed by the radio signal propagation path. In addition, using the semantic information associated with these cells, we are able to determine which analytical model to apply in order to precisely compute the path loss.

We have shown a tool that leverages the enriched description of the IVGE and computes the radio signal attenuation due to buildings, foliage and field obstructions. However, other phenomena can also degrade the radio signal transmission. Examples of such phenomena include transmitter power, receiver sensitivity, and radio signal's absorption, reflection, and scattering from interaction with features on or near its transmission path. These phenomena should be taken into account when computing the radio signal attenuation in order to predict anticipated communications network connectivity and performance data. To this end, the agent models representing the transmitters and receivers antennae should be improved in order to take into account the antenna sensitivity. Moreover, the computation of the signal propagation should also take into account absorption, reflection and scattering phenomena.

In the future, we propose to extend our tool in order to integrate an advanced ray-tracing process [10], which combines both the geometric optics and the geometric theory of diffraction (GTD). Moreover, we propose to include the uniform theory of diffraction (UTD) [8] extension to GTD which removes the inaccuracies close to the incident and reflection boundaries. Since we already carried out in the field measurements of radio signal attenuation within the metropolitan area of Windsor, Ontario, Canada [6], the next step is to build the IVGE representing this geographic area using GIS data and reproduce using our agent-based simulation tool these measurements for validation and calibration

 ISBN: 978-1-61208-118-2

purposes of our radio signal attenuation model.

To conclude, our geometrically-precise and semantically-enhanced IVGE enables us to provide wireless network planners with a tool for the analysis of the communications' attenuation. In contrast with mathematical models which only approximate the radio signal attenuation based on a coarse-grained qualification of the geographic environment: *urban*, *suburban* and *rural*, we compute more precisely the radio signal propagation path and qualify obstructions in order to apply the appropriate analytical model.

ACKNOWLEDGMENT

This research is supported by the Canadian Network of Centres of Excellence in Geomatics (GEOIDE), and by the Natural Sciences and Engineering Research Council (NSERC) of Canada. M. Mekni benefited from a PhD scholarship granted by FQRNT (Fonds Québécois de la Recherche sur la Nature et les Technologies). Authors thank Dr Phil Graniero, Univ. of Windsor for his advice during this research.

REFERENCES

[1] W. Ali and B. Moulin. 2D-3D multiagent geosimulation with knowledge-based agents of customers' shopping behavior in a shopping mall. In *Spatial Information Theory*, pages 445–458. Elsevier, 2005.

[2] I. Benenson and P. Torrens. *Geosimulation: Automata-Based Modeling of Urban Phenomena*. John Wiley and Sons Inc., 2004.

[3] G. Comparetto, J. Schwartz, N. Schult, and J. Marshall. A communications analysis tool set that accounts for the attenuation due to foliage, buildings, and ground effects. In *MILCOM 2003: Proceedings of the Military Communications Conference*, volume 1, pages 1 – 5, Boston, MA, USA, October 2003.

[4] S. Donikian and S. Paris. Towards embodied and situated virtual humans. In *Motion in Games*, pages 51–62, 2008.

[5] N. Farenc, R. Boulic, and D. Thalmann. An informed environment dedicated to the simulation of virtual humans in urban context. In P. Brunet and R. Scopigno, editors, *Computer Graphics Forum (Eurographics '99)*, volume 18(3), pages 309–318. The Eurographics Association and Blackwell Publishers, 1999.

[6] P. Graniero. A spatial analysis of gps availability and radio data transmission reliability for real-time, wireless gis updating in urban environments. Technical report, MEMF Lab, University of Windsor, Dept. of Earth Sciences, 2004.

[7] P. Graniero and H. Miller. Real-time, wireless field data acquisition for spatial data infrastructures. In *GeoTec Event 2003*, Vancouver BC, Canada, March 2003.

[8] R. Kouyouijian and P. .Pathak. A uniform geometric theory of diffraction for an edge in perfectly conducting surface. In *IEEE*, volume 62, pages 1448–1461, 1974.

[9] J.-E. Marvie, J. Perret, and K. Bouatouch. Remote interactive walkthrough of city models. In *Proceedings of the 11th Pacific Conference on Computer Graphics and Applications (PG'03)*, pages 389–393, Oct. 2003.

[10] R. Matschek. A geometrical optics and uniform theory of diffraction based ray tracing optimisation by a genetic algorithm. *Comptes rendus de physique*, 6(6):595–603, July 2005.

[11] M. Mekni. *Automated Generation of Geometrically-Precise and Semantically-Informed Virtual Geographic Environments Populated with Spatially-Reasoning Agents*. Dissertation.com, August 2010.

[12] S. Paris, M. Mekni, and B. Moulin. Informed virtual geographic environments: an accurate topological approach. In *The International Conference on Advanced Geographic Information Systems & Web Services (GEOWS)*, pages 1 – 6. IEEE Computer Society Press, 2009.

[13] S. Rodriguez, V. Hilaire, S. Galland, and A. Koukam. Holonic modeling of environments for situated multi-agent systems. In *Environments for Multi-Agent Systems II*, pages 18–31. 2006.

[14] J. Schwartz, G. Comparetto, and J. Marshall. The communications resource planning tool. In *Military Communications Conference, 2003. MILCOM 2003. IEEE*, volume 1, pages 227–230, Boston, MA, USA, October 2003.

[15] S. Seidal and T. Rappaport. Site-specific propagation prediction for wireless in-building personal communication system design. *IEEE Transactions on Vehicular Technology*, 43(4):879–891, 1994.

[16] W. Shao and D. Terzopoulos. Environmental modeling for autonomous virtual pedestrians. *Digital Human Modeling for Design and Engineering Symposium*, 2005.

[17] G. Thomas and S. Donikian. Virtual humans animation in informed urban environments. *Computer Animation 2000*, pages 112–119, 2000.

[18] M. Weissberger. An initial critical summary of models for predicting the attenuation of radio waves by trees. In *Final Report Electromagnetic Compatibility Analysis Center, Annapolis, MD.*, Annapolis, Md, July 1982.

[19] S. Y. Willassen. A method for implementing mobile station location in GSM. Master's thesis, Norwegian University of Techniques and Sciences, 1998.

[20] J. Zhu, J. Gong, H. Lin, W. Li, J. Zhang, and X. Wu. Spatial analysis services in virtual geographic environment based on grid technologies. *MIPPR 2005: Geospatial Information, Data Mining, and Applications*, 6045(1):604–615, 2005.

 ISBN: 978-1-61208-118-2

Determining the Geographical Origin of a Serial Offender Considering the Temporal Uncertainty of the Recorded Crime Data

Marie Trotta, Benoît Bidaine, Jean-Paul Donnay
Geomatics Unit
University of Liège, Belgium
{Marie.Trotta, B.Bidaine, JP.Donnay}@ulg.ac.be

***Abstract*—Since the days the investigating officers used "pin maps" to locate and to think about crime events, crime mapping has become widespread thanks to spatial analysis mainly supplied by GIS-like software. In particular these methods suit well to geographic profiling devoted to crime series characterised by a single offender and hence limited space and time variability. Although spatial techniques are now regularly performed to delineate an offender's area of residence, the temporal dimension is underemployed due to the wider uncertainty of time records. This paper proposes a methodology based on a least-squares adjustment in order to cope with this temporal issue for determining the most probable offender's residence. Moreover, a chi-square test is described to check the significance of the solutions suggested by the method. The process is carried out on the real road network which has been discretised (rasterised) for computing convenience. Three simulations show the validity of the reasoning. Finally the main time and speed assumptions introduced in the model are discussed paving the way for further research.**

***Keywords*-crime mapping; geographic profiling; temporal data quality; least-squares adjustment; raster diffusion process; error propagation**

I. INTRODUCTION

Among the domains where the quality of spatial and temporal data especially affects operational performance analysis, crime mapping is probably one of the most intriguing to the general public. Its premises date back to the 19th century, but the analysis offered by GIS allowed geographic profiling to expand only in recent decades. Geographic profiling is defined as a methodology of investigation that uses the locations of a series of connected crimes to determine the criminal's most probable area of residence [1]. In addition to psychologist profiling, it is nowadays widely used in serious crime investigations.

The temporal dimension of geographic profiling has been underexploited while recent criminal literature underlines the importance of simultaneously addressing both spatial and temporal aspects of crimes (several references in [2]). Uncertainty in temporal data largely explains their underutilization. The failure to capture temporal details by the police [3], the practical and/or psychological inability of the victim to specify the time of a crime, and the variability of the offenders' behaviour are the main causes of the low reliability of temporal data. If temporal information is processed in conjunction with spatial information, errors may appear on space (sought-after criminal's place of residence, crime location), on time (starting time of the criminal from his/her residence, time of the crime) and on speed (travel from criminal's residence to crime site).

This paper precisely deals with the modelling of the inaccuracies of the recorded times of crime for providing geographic profiling with a new method exploiting spatio-temporal convergence. In order to address this issue, the paper is organised as follows. The context, Section II, describes the situation in which the offender's anchor site can be determined thanks to temporal information. It also specifies the temporal, behavioural and spatial assumptions required to limit the variation of spatial and temporal parameters. In Section III, we present the basic isotropic methodology used to locate the criminal's most probable area of residence. Then we analyse the least-squares solution for the generalization of the diffusion process on an anisotropic space. Subsequently a chi-square test is proposed in order to assess the validity of the improved method. The reasoning is then illustrated in Section IV by simulation processes generating variances in the data sets.

The consequences of the assumptions introduced in the previous process are discussed in Section V and methodological solutions are considered in order to raise these hypotheses. Finally, the conclusion in Section VI summarises the bringing-in of this article.

II. CONTEXT

The context chosen for this research is dictated by real facts relating to multiple rapes perpetrated by the same author, at different dates and for several years within a reduced spatial and temporal variability: crimes spread across an area limited to a radius of a few tens of kilometres, and are perpetrated in a winter time slot ranging between 7.00 and 7.30 AM. The narrow time slot during which the crimes were committed suggests a strong but likely assumption, i.e. a constant departure time from a single anchor point. By modelling the travel time from each crime site, it should then be possible to identify, back in time, the location with similar

 ISBN: 978-1-61208-118-2

departure time. This one is assumed to be the offender's anchor point.

Concerning the offender's spatial behaviours, two kinds of scenarii could be explored. According to the criminal profile and the crime type, the offender can take advantage of an opportunity during his daily activities (theory of routine activity in environmental criminology [4]) or, conversely, he prepares the crime beforehand by a precise location.

The first behaviour can use several types of travel between the criminal's place of residence and the site of crime (circuit or zigzag for example). The second behaviour, on the other hand, strongly favours the direct travel between his residence and the site of crime i.e. the use of the shortest path on the road network. In our context, the crime locations and the times they were committed suggest that the criminal has made a very specific location prior to each crime corresponding to the second behaviour. It is therefore possible to favour a "star" pattern in the journeys between the criminal's anchoring site and the sites of crimes.

Regarding the spatial assumptions, the extending area of crime is too large to consider a pedestrian behaviour. Assuming car as mode of transportation, the journeys-to-crime are limited to the road network and only crimes carried in its vicinity are considered. All the points of the network are then candidates to host the offender's anchor point.

III. Determining the geographical origin of a serial offender

Our research question is defined as follows: is it possible to determine the geographical origin of a serial offender taking into account the temporal uncertainty of the crime data? The analysis is based on the temporal, behavioural and spatial assumptions described in the context.

This section analyses firstly the research question with a simplified reasoning. Then it introduces gradually the complexity of real cases in order to clearly identify the various issues relating to the problem.

A. Isotropic diffusion case

1) Definition of the problem: n criminal events occurred at places with known coordinates (x_i, y_i) and at recorded times t_i (with $i = 1$ to n). The purpose of the analysis is to determine the offender's origin, assumed to be his/her residence, and his/her constant time of departure. The problem contains three unknowns: x and y the coordinates of the offender's residence and t, the time of departure from this residence. In the specific case of isotropic diffusion and constant velocity v, the problem is analytically formulated as follows:

$$t + \frac{\sqrt{(x_i - x)^2 + (y_i - y)^2}}{v} = t_i \tag{1}$$

In the case described above, we may identify a solution for this problem according to the least-squares adjustment (LSA) in order to take into account the variability of the t_i. This variability has multiple sources: a bad estimation given by the victims, some waiting time before acting for the offender, small variations in his time of departure.

2) Least-squares formulation: Residuals ν_i were added to the n observation equations:

$$t + \frac{\sqrt{(x_i - x)^2 + (y_i - y)^2}}{v} = t_i + \nu_i \tag{2}$$

According to least-squares theory [5], [6], these equations are linearised:

$$t + \frac{D_i^o}{v} + \frac{(x^o - x_i)}{vD_i^o}\,(x - x^o) + \frac{(y^o - y_i)}{vD_i^o}\,(y - y^o) = t_i + \nu_i \tag{3}$$

where (x^o, y^o) is an approximation of the residence coordinates and D_i^o the corresponding distance to the crime i location. These n equations can be written in matrix form:

$$A\ \underline{x} - \underline{\nu} + \underline{w} = \underline{0} \tag{4}$$

where

$$A = \begin{pmatrix} 1 & \frac{x^o - x_1}{vD_1^o} & \frac{y^o - y_1}{vD_1^o} \\ \vdots & \vdots & \vdots \\ 1 & \frac{x^o - x_n}{vD_n^o} & \frac{y^o - y_n}{vD_n^o} \end{pmatrix} \tag{5}$$

$$\underline{x} = \begin{pmatrix} t \\ x - x^o \\ y - y^o \end{pmatrix} \tag{6}$$

$$\underline{w} = \begin{pmatrix} \frac{D_1^o}{v} - t_1 \\ \vdots \\ \frac{D_n^o}{v} - t_n \end{pmatrix} \tag{7}$$

The Jacobian matrix A contains the equations partial derivatives with respect to the unknowns. The vectors $\underline{x}$, $\underline{\nu}$ and $\underline{w}$ gather respectively the unknowns, the residuals and the independent terms.

The least-squares solution is then given by

$$A^T A\ \underline{\hat{x}} = -A^T \underline{\hat{w}} \tag{8}$$

$$\underline{\hat{x}} = -(A^T A)^{-1} A^T \underline{\hat{w}} \tag{9}$$

where $\underline{\hat{x}}$ is the estimation of x constituting the solution in the least-squares sense .

By definition, this solution minimises the sum of the squares of the residuals (SSR): $\underline{\nu}^T \underline{\nu}\ min$

B. Diffusion on the network

The previous section shows that the simplified problem can be solved with the LSA. However, we assumed an isotropic diffusion on a continuous space. In real situations the road network conditions the path followed by the offender. Therefore we need to generalise the problem to

 ISBN: 978-1-61208-118-2

any diffusion process, which from a practical standpoint, is treated by a discretisation of space. Equation 2 becomes:

$$t + d_i(x, y) = t_i + \nu_i \tag{10}$$

where $d_i(x, y)$ are the travel times or the delays between the origin (x, y) and the crime i. The travel time or equivalently the potential starting time at any point of the network $t_{o,i}(x, y)$ is estimated numerically. The solution will be obtained semi-analytically. Let us denote by $\bar{t}(x, y)$ the mean of the potential starting times between homologous locations (x,y):

$$\bar{t}(x, y) = \frac{\sum_{i=1}^{n} t_{o,i}(x, y)}{n} \tag{11}$$

then the least-squares solution is the triplet $(x, y, \bar{t}(x, y))$ minimising the SSR:

$$\underline{\nu}^T \underline{\nu} = \sum_{i=1}^{n} (\bar{t}(x, y) - t_{o,i}(x, y))^2 \ min \tag{12}$$

C. Statistical validation

According to least-squares hypotheses, residuals follow a normal distribution. Consequently, the SSR follows a χ^2 distribution with $n - p$ degrees of freedom (with n the number of observations and p, the number of unknowns). Therefore the following χ^2 test was built to determine the upper bound of the SSR below, which the area can potentially contain the offender's residence.

$$\underline{\nu}^T \underline{\nu} < \ (n - p)\ \sigma^2 \chi^2_{n-p} \tag{13}$$

σ^2 is the a-priori variance chosen according to the uncertainty attributed to the recorded t_i. In first approach, we postulate that the uncertainty is similar for all the observed times (t_i). Indeed, this uncertainty influences the trust we attribute to the identified solution.

IV. Simulations

We chose to work in raster mode as all the pixels of the road network could potentially be candidates for the anchor point. Indeed, the raster mode is suitable to represent a spatially continuous phenomenon. Besides, this mode is compatible with parallel developments explained in the perspectives. Three simulations, implemented on ArcGis 9.3 (Spatial Analyst) are performed to illustrate the reasoning with the t_i obtained from a randomly chosen residence and starting time.

The first simulation considers the t_i without uncertainty while the two others introduce two different precisions of 2 and 5 minutes on the t_i.

Spatial data concerning the road network and the sites of crimes are available with high precision (5 m). By contrast, there is no information on the network load and therefore on the varying vehicle speeds on the different parts of the network at crime times. In order to isolate the temporal uncertainty coming from the times of crime, a constant speed of 50 km / h was set for this application. It is worth noting that a constant velocity could correspond to a motorist looking for potential targets as well as, for a lower speed, to a pedestrian movement.

A. Evaluation of times corresponding to the crime events (t_i)

The first step of the simulation is to determine four plausible times of crime at four distinct locations respecting the assumption of a constant departure time from an unique location.

- We first arbitrarily choose a location for the offender's residence and a unique time of departure corresponding to 7 AM.
- We build a cost surface corresponding to the crossing time at an assumed constant speed of 50 km/h. The surface is generated in raster mode with a 10-m pixel resolution. A cost of 0.072 is assigned to every pixel of the rasterised network, corresponding to the time required to travel one meter at this speed of 50km/h.
- The cost distance from the offender's residence d is calculated in each cell of the network as a sum of costs per cell c_j. We computed the cost distance using the function "Cost Distance" in ArcGis based on the following algorithm [7].

$$c_j = c * r \tag{14}$$

with c the unit cost and r the cell resolution

$$a_j = \frac{c_j + c_{j+1}}{2} * k \tag{15}$$

where a_j is the accumulative cost to move from cell j to cell $j + 1$ (cf. Fig. 1) and

$$k = \begin{cases} 1 & \text{if the movement is horizontal or vertical} \\ \sqrt{2} & \text{if the movement is diagonal} \end{cases} \tag{16}$$

$$d = \sum_{j=1}^{m} a_j \tag{17}$$

where j is the current pixel on the way between the origin and the destination and m is the number of pixels on this same way.

- The departure time is added to the calculated cost distance to generate the crossing time of each cell in the network.
- Four crime sites are randomly selected on the network and their crossing time is considered as the crime time ($n = 4$).

 ISBN: 978-1-61208-118-2

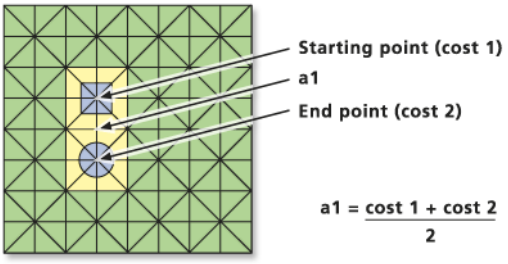


Figure 1. Horizontal and verical node calculations. [7]

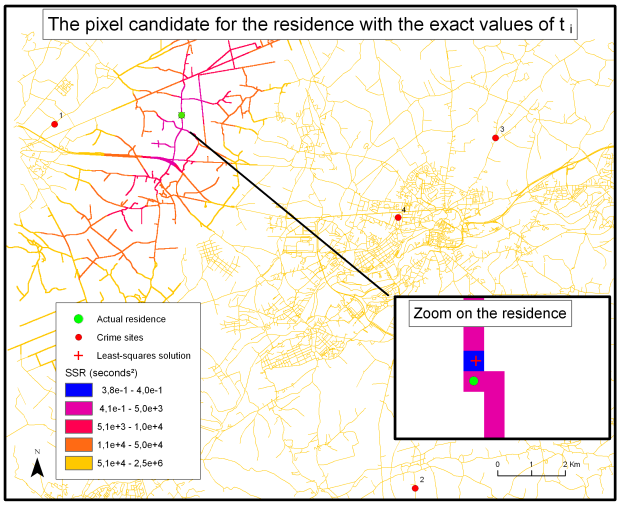


Figure 2. The almost-overlap of the actual residence and the least-squares solution, minimizing the sum of the suared residuals (SSR), illustrates the validity of the least-squares adjustment.

B. LSA for the exact values of t_i

The potential starting time for each observation t_i in every cell of the network $t_{o,i}$ is evaluated by reverse processing (regressive time from the crime sites through the road network using the same cost). This step generates four images that will be used as input for the LSA.

As explained in Section III-B, the pixel average of these images is calculated. This corresponds to the average starting time for the analysed crimes. The residuals consist in the differences between this average $\bar{t}$ and each observation $t_{o,i}$ computed at every pixel of the network. The least-squares solution of the system is the cell that minimises:

$$\sum_{i=1}^{n}(\bar{t} - t_{o,i})^2 \qquad (18)$$

and should ideally be equal to 0.

The result is presented in Fig. 2. The obtained minimum is slightly different from 0 and located in the cell just next to the one containing the arbitrary chosen residence. The result can still be considered as valuable despite this difference explained by the algorithm used to calculate the

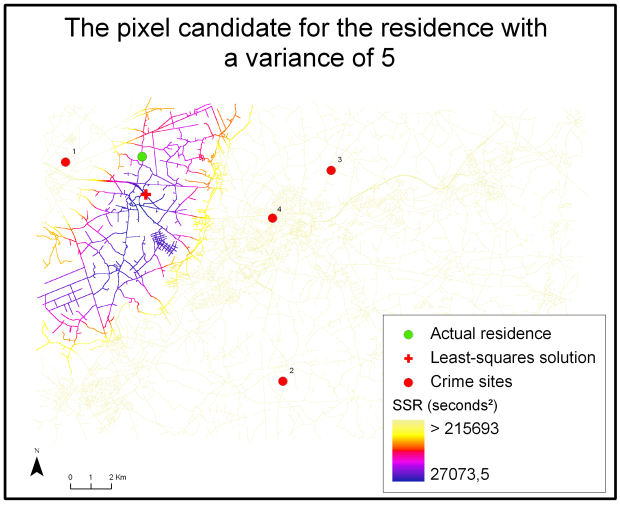


Figure 3. The offender's residence is located inside the area delineated by the χ^2 threshold on the sum of the squared residuals (SSR).

cost distance. Indeed, the cost is calculated from the cell next to the one containing the residence and its value depends on the precision of the cell resolution.

C. Introduction of the uncertainty on the t_i

We choose two a-priori variances of 5 and 10 corresponding respectively to uncertainties of 2'15" and 3'10". The new values of t_i consistent with these variances, respectively $t_{i,2min}$ and $t_{i,5min}$, are presented in Tab. I. The potential starting time in every cell of the network is then updated using the methodology previously described in order to evaluate in each cell the new value of SSR. Figures 3 and 4 illustrate the result of the test described in Equation 13 using a probability level of 95 %. The least SSR value, the least-squares solution, does not correspond to the arbitrary offender's residence because of the introduced uncertainties. Nevertheless it is worth noting that this place remains located inside the area delineated by the thresholded value of the χ^2 test.

In addition, the search for the criminal can be prioritised in this area.

V. Discussion of the Assumptions and Perspectives

The previous analyses are based on several assumptions that are not necessarily encountered in real life.

Firstly, it assumes an offender's constant starting time. This is certainly the most restrictive hypothesis as it renders useless any attempt to solve crime series where the crime locations are close to each other while the crime times are very different. Therefore, other hypotheses than a constant departure time of the offender have to be considered. The minimization of the variances of the journey times from

 ISBN: 978-1-61208-118-2

Table I
SIMULATION PARAMETERS

Crime event	Distance (km)	Travel time	Exact t_i	t_i with a variance of 5	t_i with a variance of 10
1	5.11	6'08"	07:06:08	07:08:22	07:10:08
2	15.89	19'04"	07:19:03	07:16:37	07:21:31
3	12.31	14'46"	07:14:46	07:16:46	07:11:46
4	8.50	10'12"	07:10:12	07:07:58	07:07:12
Value of the upper bounds calculated from the χ^2 (seconds^2)				215693	431385

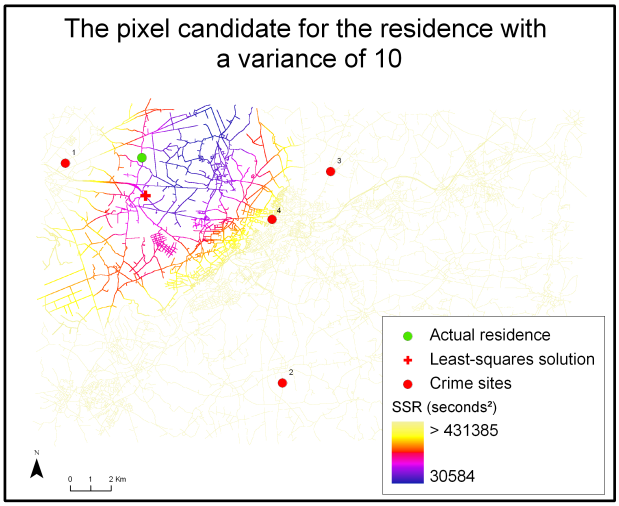


Figure 4. The delineated area widens and moves with the change of temporal uncertainty.

the offender's anchor point could be analysed with a quite similar methodology.

Secondly, an identical time uncertainty is attached to every observation. In practice some events may be recorded with better precision than others: a witness can corroborate the information given by the victim; the time of the event can be constrained by the victim's activity schedule; etc. In order to consider such a variable time uncertainty, a solution could be obtained thanks to a weighted LSA [5]. Indeed the relative confidence attributed to each crime time could be computed as the ratio of the a priori variance and the corresponding time variance. These weights introduced in the resolution process would then modify the estimation of the unknowns as well as the SSR analysed in the χ^2 test.

A third assumption deals with the speed supposed to be constant going through the road network. If this assumption can be considered acceptable for pedestrian journeys, it is not the case for an offender travelling by car. Car cruising speed varies considerably on a network according to a number of parameters depending notably on the spatial environment (urban street, rural road, etc.) and on temporal conditions (time of the day, day of week, etc.). Most of these parameters are well identified in the traffic engineering literature [8]–[10] and may then be exploited by our methodology with very little changes. However the speed at a precise place and moment can only be approximated by the dedicated algorithms (e.g., multi-agent system) [11], [12]. Consequently the errors inherent to speed are added to the time uncertainties on the events and they are merged in the residual values provided by the LSA.

Besides, crimes are generally not committed on the road as considered in this application, but either in discrete locations in the vicinity of the road where the victims are driven, or in the victims'home. The path covered by foot by the criminal on and out of the road can reach tens to hundreds of meters. A research trying to model this kind of trips is under progress, aiming to achieve a cost surface using raster images of land uses at a metric or decametric resolution.

The fact of raising assumptions modifies the assessment of the total time error. The method presented herewith should therefore be part of a more comprehensive approach to error propagation. In this respect the use of a modified Monte Carlo algorithm to train different types of Bayesian neural networks and to estimate uncertainty limits [13], is considered.

VI. CONCLUSION

This study develops a methodology for determining the most probable area for an offender's residence assuming a constant starting time. The process is performed on a rasterised road network crossed at constant speed and the method, which is based on a LSA, is able to include the uncertainty affecting the recorded times of the crimes. A χ^2 test also described herewith allows to check the significance of the value presented by the residence locations (pixels) suggested by the method.

This method allows to test a hypothesis of constant starting time for the offender's spatial behaviour at a constant travelling speed. A geographic profile can only be built on eliminating progressively some hypothesis concerning this behaviour.

The analysis is supported by three simulations: the first one assumes exact values for the crime times, while the others introduce a variability in the crime times corresponding to variances of 5 and 10 respectively. The simulations show that the identified solutions lie in the vicinity of the correct location in a way adequately described by the χ^2 statistic.

 ISBN: 978-1-61208-118-2

ACKNOWLEDGMENT

The researches achieved by M. Trotta and B. Bidaine are funded under F.R.S.-FNRS fellowships (Belgian National Fund for Scientific Research).

REFERENCES

[1] K. Rossmo, *Geographic profiling*. Boca Raton.: CRC Press, 2000.

[2] T. Grubesic and E. Mack, "Spatio-temporal interaction of urban crime," *Journal of Quantitative Criminology*, vol. 24, no. 3, pp. 285–306, 2008.

[3] P. J. Brantingham and P. L. Brantingham, "Anticipating the displacement of crime using the principles of environmental criminology," pp. 119–148, 2003.

[4] P. Brantingham and P. Brantingham, "Notes on the geometry of crime," in *Environmental Criminology*, B. P.L. and B. P.J., Eds. Beverly Hills: Sage, 1981, pp. 27–54.

[5] C. Ghilani, *Adjustment Computations: Spatial Data Analysis*, 2nd ed. John Wiley & Sons, 2010.

[6] P. Sillard, *Estimation par moindres carrés*, ser. Collection ENSG . IGN. Paris: Hermès Sciences Publications, 2001.

[7] ESRI, "Arcgis resource center. how cost distance tools work," 1999-2010. [Online]. Available: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html. Consulted on the 23th October 2010

[8] W. Homburger, J. Hall, W. Reilly, and E. Sullivan, *Fundamentals of Traffic Engineering*, 16th ed. University of Berkeley, Institute for Transportation Studies, 2007.

[9] H. Miller and S.-L. Shaw, *Geographic Information Systems for Transportation*. Oxford University Press, 2001.

[10] J.-C. Thill, *Geographic Information Systems in Transportation Research.* Pergamon, 2000.

[11] E. Groff, "Characterinzing the spatio-temporal aspects of routine activities and the geographic distribution of street robbery," in *Artificial Crime Analysis System. Using Computer Simulation and Geographic Information Systems*, L. Liu and J. Eck, Eds. New York: Information Science Reference, 2008, pp. 226–251.

[12] F. Balbo and S. Pinson, "Dynamic modeling of a disturbance in a multi-agent system for traffic regulation," *Decision Support Systems*, vol. 41, no. 1, pp. 131–146, 2005.

[13] F. L. Xuesong Zhang, R. Srinivasan, and M. V. Liew, "Estimating uncertainty of streamflow simulation using bayesian neural networks," *Water Resour. Res.*, vol. 45, no. 2, p. 16, 2009. [Online]. Available: http://dx.doi.org/10.1029/2008WR007030

 ISBN: 978-1-61208-118-2

Water Area Extraction from RGB Aerophotograph Based on Chromatic and Textural Analysis

Meng Zhao Huazhe Shang Wenchao Huang Lizhi Zou Yongjun Zhang
School of Remote Sensing and Information Engineering,
Wuhan University
Wuhan, China
zmatwhu@yahoo.cn 754773310@qq.com 423036600@qq.com 867750650@qq.com yongjun_zhang@sina.com

***Abstract*—When triangulating RGB aerophotograph, if automatically and randomly selected matching pass points unfortunately locates into water areas, these points, limited by their inaccuracy, will decrease the precision of triangulation. Therefore, extraction of water area beforehand is conducive to eliminate water falling pass points and guarantee the quality of aerotriangulation. A new methodology to extract water area from RGB aerophotograph is put forth in this paper. Procedure initiates by segmenting the whole aerophotograph into homogenous and united segments. Subsequently, compute chromatic and textural features of every segment and compare each segment's features to sampled water segments' features. Finally extract those segments whose chromatic and textural features are similar to sampled segments' as water areas. This methodology has a relatively obvious merit in effectiveness and generality.**

Keywords—CIELAB; chromatic analysis; textual analysis; watershed segmentation; ISODATA

I. INTRODUCTION

Aerotriangulation is a key step in aerophotogrammetry. The basic goal of aerotriangulation is to compute all elements of exterior orientation and pass points of a region by block adjustment. The accuracy of matching points on aerophotograph directly relates to the precision of block adjustments and the final production of aerotriangulation. However, water areas, influenced by wind force and gravity, are usually in irregular motion. If automatically and randomly selected matching pass points (a few pass points selected beforehand to compute the exterior orientation elements) unfortunately falls into water areas, these points, limited by their inaccuracy, will decrease the precision of triangulation. Therefore, extraction of water areas beforehand is conducive to eliminate water falling pass points and guarantee the quality of aerotriangulation. In this paper, we concentrate on extracting water areas from RGB aerophotograph and mark them.

Multispectral aerophotograph and remote sensing image can synthesize information of different spectrums to extract water region. J. Deng operated different bands of SPOT-5 images to extract water area [1]. H. Xu modified MNDWI value to extract water area from multispectral remotely sensed data [2]. In contrast to multispectral data, RGB aerophotograph (every pixel of the image is consisted of red, green and blue values) has only three bands and little extra spectrum information. Consequently, until recently, no effective methodology has been proposed to extract water area from RGB aerophotograph. In order to supply such a gap, we put forth a methodology in this paper to extract water area from RGB aerophotograph.

With the enhancement of image resolution, aerophotographs contain more abundant space information as well as geometric and textual features, which create a favorable condition for extracting objects via chromatic and textual analysis. W. Ma and B. S. Manjunath had used textural analysis to create a texture thesaurus for browsing large aerophotographs and achieved relatively good retrieval effect [3]. Other trials include W. Niblack's querying image using color, texture and shape [4]. All these researches manifest a fact that chromatic and textural analyses are effective in extracting and retrieving objects.

Enlightened by chromatic and textural analysis, we propose a new series of procedures to extract water areas from RGB aerophotographs. Procedure initiates by segmenting the whole aerophotograph into homogenous and united regions. Subsequently, compute chromatic and textural features of every region and then compare each region's feature to sampled water regions' features. Finally extract those segments whose chromatic and textural features are similar to sampled segments' as water areas. This methodology has a relatively obvious merit in effectiveness and generality.

Water areas on aerophotographs are usually homogenous and united; therefore water regions after segmentation always are homogenous and united. We have no need to care about the accuracy in segmentation of other objects because they contribute little to water extraction. We choose watershed segmentation algorithm (Section Ⅲ) because it can help attain satisfactory segmentation result. The homogenous and united characteristic of water region also provides pleasing condition to conduct chromatic analysis (Section Ⅱ) and textural analysis (Section Ⅳ). Chromatic and textural features of every region will be added to its own feature vector (Section Ⅴ). Water appears differently on photographs. In order to automatically recognize water areas, we need to sample water regions of all appearances beforehand and classify them. ISODATA algorithm (Section Ⅴ) is employed to classify those samples into several categories. If the distance is within a threshold between a certain region's feature vector and any one of those

 ISBN: 978-1-61208-118-2

categories', then this segment will be identified and extracted as water area.

RGB aerophotograph consists of three channels (Red, Green, and Blue) and they are closely pertinent to each other. We need to integrate them into one channel in order to conduct image segmentation and textural analysis. In this paper, we utilize CIELAB color space to integrate those three channels, which decreases their pertinence as well as provides prerequisites for image segmentation and textural analysis later on.

Any color is unable to be both green and red or both blue and yellow. This principle prompts the CIELAB color space. CIELAB color space inherits the XYZ standard color system. CIELAB indicates distinctions between light and dark, red and green, and blue and yellow by using three axes: L*, a*, b*. The central vertical axis represents lightness (L*) whose value ranges from 0 (black) to 100 (white). On each axis the value runs from positive to negative. On a-a' axis, positive values indicate amounts of red while negative indicate amounts of green. On the b-b' axis, yellow is positive and blue is negative. For both axes, zero is neutral [5].

"Texture" refers to the arrangement or characteristic of the constituent elements of anything. A texture feature is a value, computed from the image of a region, which quantifies gray-level variation within the region. According to Kenneth [6], a texture feature value is irrelevant to region's position, orientation, size, shape and brightness (average gray level). Therefore, the shape and size of segmented area do not interfere with the result of chromatic and textural analysis.

According to Julesz [7] and his deduction, human texture discrimination is based on the second order statistic of image intensities, which gave rise to the emergence of a popular textural descriptor: co-occurrence matrix. A co-occurrence matrix counts the exact times of different grey level pairs of pixels, separated by a certain distance (one pixel, two pixels, etc.). The (i, j) element of the co-occurrence matrix P for a certain region is the number of times, divided by M, that gray levels i and j occur in two pixels separated by that distance and lying along that direction in the region, where M is the number of pixel pairs contributing to P. The P matrix is N by N, where the gray scale has N shades of gray [6].

Watershed segmentation is an approach based on the concept of morphological watersheds. In such a "topographic" interpretation, every pixel's gray level denotes its "height". For a particular regional minimum, the set of points at which a drop of water, if placed at the location of any of those points, would fall with certainty to a single minimum is called the catchment basin. And the points at which a water drop would be equally likely to fall to more than one such minimum constitute crest lines on the topographic surface and termed watershed lines. Assume that a hole is punched in each regional minimum and that the entire topography is flooded from below by letting water rise through the holes at a same rate. When the rising water from different catchment basins is about to merge, a dam is built to prevent the merging. The flooding will finally reach a stage when only the tops of the dams are visible above the water lines. These dam boundaries correspond to the divide lines of the watersheds. Therefore, they are boundaries extracted by a water shed segmentation algorithm [7].

II. CIELAB COLOR SPACE AND CHROMATIC ANALYSIS

There are two steps to transform RGB color space to CIELAB color space [5]. The first step is to apply the following matrix to convert RGB color space to CIEXYZ color space.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

Subsequently, (X^*, Y^*, Z^*) need to be calculated.

When $X / X_n, Y / Y_n, Z / Z_n > 0.008856$,

$$X = \sqrt[3]{X / X_n} \quad (2)$$

$$Y = \sqrt[3]{Y / Y_n} \quad (3)$$

$$Z = \sqrt[3]{Z / Z_n} \quad (4)$$

When $X / X_n, Y / Y_n, Z / Z_n \le 0.008856$,

$$X^* = 7.787 \cdot (X / X_n) + 0.138 \quad (5)$$

$$Y^* = 7.787 \cdot (Y / Y_n) + 0.138 \quad (6)$$

$$Z^* = 7.787 \cdot (Z / Z_n) + 0.138 \quad (7)$$

(X_n, Y_n, Z_n) = (0.312779, 0.329184, 0.358037) represents the white reference point, which indicates a completely matte white body.

Then L*, a*, b* can be calculated through following equations:

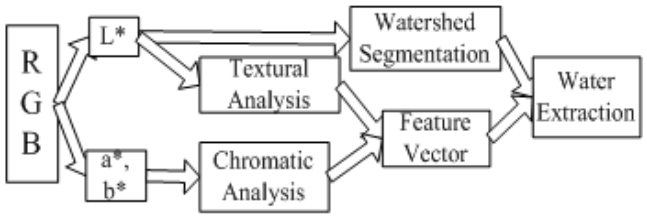
$$L = 116 \cdot Y^* - 16 \quad (8)$$

$$a^* = 500 \cdot (X^* - Y^*) \quad (9)$$

$$b^* = 200 \cdot (Y^* - Z^*) \quad (10)$$

The application of CIELAB color space in this paper can be briefly illustrated using diagram Ⅰ.

DIAGRAM Ⅰ APPLICATION OF CIELAB COLOR SPACE



The L* component of CIELAB color space can be used in watershed segmentation (Section Ⅲ). L* component can also be used in textual analysis (Section Ⅳ). We use equations mentioned above to convert every pixel on an

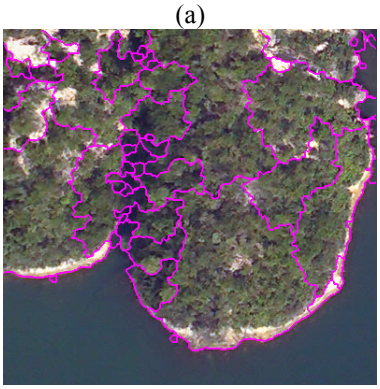
aerophotograph from RGB color space to CIELAB color space. Then form a gray image by using the L* component of every pixel's CIELAB color space. Although the gray photograph is no longer quantified 256 gray levels, L* component brings unexpected results in chromatic analysis and textural analysis later on.

Segmented areas usually have similar color space and we can use the average a*, b* of all pixels in a region to represent its chromatic feature. Experiments have demonstrated that chromatic feature between non-water areas and water areas is obvious, which means a*, b* can be employed to chromatically describe a region. We utilize a* and b* as parameters to describe a segmented region chromatically in feature vector (Section Ⅴ). If regions have similar chromatic feature with sampled water areas, they are possibly water areas we want to extract.

III. IMAGE SEGMENTATION

In our methodology, we replace the 256 gray level with L* component of CIELAB color space. Using this replacement, we could form a gray picture by synthesizing three channels of RGB aerophotographs. Compared to simply using average gray level of three channels to form a gray photograph, the usage of L* component can decrease fragments after segmentation and attain a relatively clear and accurate edge.

(a)



(b)

Figure 1. Comparison of Segmentation: (a) is attained by using the average gray level of RGB to form a gray picture. (b) is attained by using L* component of CIELAB color space to form a gray picture

In Fig. 1, (b) shows a correct and clear boundary between water area and non-water area, while result (a) displays an inaccurate boundary. This manifests effectiveness and correctness of the L* component in segmenting RGB aerophotographs.

IV. TEXTURAL ANALYSIS

In our methodology, we replace traditional 256 scale gray with the L* component of CIELAB color space. We attain two advantages by this substitute. First of all, the RGB components of aerophotograph has been integrated into one L* component, which ranges from 0 to 100; in addition, the dimension of co-occurrence has been decreased from 256*256 to 101*101. Traditionally, we reduce quantization level of the input data (e.g., from 8-bit data with values from 0-255 to 5-bit data with values from 0 to 31) when creating co-occurrence matrices so that the matrices will not become too large. But this method usually causes texture information loss, especially detailed information. Taking advantage of the L* component of CIELAB color space, the co-occurrence matrices will decrease its dimension with relatively little textural information loss.

Haralick proposed 14 features based on co-occurrence matrices [8]. The angular second-moment feature (ASM) is a measure of homogeneity of the image.

$$\mathrm{ASM} = \sum_{i} \sum_{j} p(i, j)^2 \qquad (11)$$

p(i, j) is the joint probability of gray pair (i, j) in co-occurrence P.

Texture is an innate property of virtually all surfaces [8]. Although water seems homogenous and with little obvious texture, texture feature can still be utilized to discriminate water area from other non-water area because water area has a bigger homogeneity than other objects.

In a homogeneous image, there are very few dominant gray-tone transitions. Hence the P matrix will have some entries of large magnitude and the ASM feature will be relatively bigger. In contrast, an image with irregular textures will have a large number of small entries and hence the ASM feature will be smaller [8]. Because water areas are usually homogeneous and united, so the ASM feature for water areas is relatively bigger than other objects. So we employ ASM as a textural descriptor.

In order to avoid texture rotating, we compute the average value of ASM in four directions (0°, 45°, 90° and 135°) to texturally describe a region. ASM is added to the feature vector (Section Ⅴ).

V. CLASSIFICATION AND WATER EXTRACTION

A. Feature Vector

A vector is a quantity that has magnitude and direction and that is commonly represented by a directed line segment whose length represents the magnitude and whose orientation in space represents the direction. Feature vector in this paper denotes that the quantity consists of a region's feature values. We define feature vector as following:

 ISBN: 978-1-61208-118-2

$$\mathbf{FV} = (\mathrm{ASM},\ a^*,\ b^*) \tag{12}$$

Different regions have different chromatic and textural features; therefore every segmented region possesses its exclusive **FV**, which represents its uniqueness among other regions. However, slight difference in **FV** exists among regions with similar textural and chromatic features, which means we could use classification algorithm to categorize regions with similar feature vectors. Every region's **FV** has to be computed for later extraction.

B. Sampling and Unsupervised Classification Using ISODATA Classification

Practical experience tells us that variances, such as contamination, fluctuation, ship transportation, etc, render water display different color and texture feature on RGB aerophotographs. Therefore, the segmented image has more than one kind of water. From the segmented image we need to sample a certain number of water regions via our practical experience. Such experience can simply tell water regions apart from other regions like forest and residence, however, it is less reliable when applied to discriminate differences within waters. To settle down this problem, we propose to use ISODATA algorithm to automatically classify sampled water regions into several categories.

The ISODATA algorithm [9], abbreviated for Iterative Self-Organizing Data Analysis Technique, is a modification of the k-means clustering algorithm. ISODATA is self-organizing because it requires relatively little human input. Therefore it is a good choice to substitute human perception to classify waters.

ISODATA algorithm begins by setting a certain amount of cluster centers for all the samples. Then classify all samples by shortest distance algorithm. Modify the centers of every cluster. If two clusters' separation distance in feature space is below a user-specified threshold, then the two cluster centers should be merged into one and reclassify all the samples. If the current iterative time is odd or the amount of clusters is half fewer than expected, then split the cluster that has the maximum subparameter of standard deviation vector and reclassify all samples. If the current iterative time is even or the amount of clusters is twice bigger than expected, then merge two clusters whose separation distance is the closest and reclassify all samples. When the iterative time has reached maximum time, then the whole algorithm stops. We have conducted quite a number of experiments and find that the result of classification is satisfactory if only the maximum standard deviation and minimum distance between cluster means are properly set (details, [9]). In our methodology, every segment's **FV** is the sample in ISODATA algorithm.

C. Water Extraction

Each sampled water region has its unique **FV**. As discussed above, similar water regions have slight difference in **FV**; therefore, ISODATA algorithm will automatically classify regions with analogous **FV** into one same class. After the completion of ISODATA algorithm, each class has a cluster center vector (**CCV**) and a standard deviation vector (**SDV**). Both have the same dimensions as **FV**. The j^{th} class's $\mathbf{CCV}_j$ and $\mathbf{SDV}_j$ can be computed out using following equations.

$$\mathbf{CCV}_j = \sum_{i=0}^{W} \frac{\mathbf{FV}_j}{W} \tag{13}$$

$$\mathbf{SDV}_j = \sum_{i=0}^{W} \sqrt{\frac{(\mathbf{FV}_j - \mathbf{CCV}_j)^2}{W^2}} \tag{14}$$

(Where W is the number of regions in j^{th} class and **FV** denotes the i^{th} region's **FV**.)

According to statistical principles, when the amount of samples in each class has reached a certain level, the distribution of segments' vectors in every class is subject to Gauss distribution. Nearly all members of the class fall within $\mathbf{CCV}_j \pm 3\cdot\mathbf{SDV}_j$. Ideally, we can sample numerous regions within a class and work out its **CCV** and **SDV**. Then compare remaining regions' feature vectors to the class center. If a region's feature vector falls within **CCV**±3·**SDV**, we identify it belonging to the class.

Unfortunately it is impossible to sample enough water regions to form Gauss distribution within a class, but we can still eliminate those non-water regions by checking whether their **FV**s locate within a threshold of a certain class's **CCV**. We define a parameter "ratio" to mark the dynamic range of each **CCV**, that is to say regions with **CCV** falling into **CCV**±ratio·**SDV** will be extracted as water area.

VI. TRIALS AND RESULTS

In order to test the feasibility of our methodology, we conduct following five trials. We define three indexes to evaluate the result of extraction. First index is "w-w", which means pixels extracted as water using methodology in this paper and perceived as water by human eyes. Second index is "w-n", which means pixels extracted as water using methodology in this paper but perceived as non-water by human eyes. Third index is "n-w", which means pixels extracted as non-water using methodology in this paper but perceived as water area by human eyes.

Trial 1: The following Fig. 2 is the water extraction effect of Fig. 1 (b). We have 9 samples of water segments. ISODATA algorithm classifies those 9 samples into 3 different types. The final extraction is attained when ratio is set at 1.1.

 ISBN: 978-1-61208-118-2

Figure 2. Extraction Result of Trial 1 by setting the ratio at 1.1

TABLE I. EVALUATION OF TRIAL 1

Fig. 2	w-w	w-n	n-w	Total Pixels
Pixels	260018	2108	18	262144
Percentage	99.2%	0.79%	0.01%	100.0%

Trial 2: The following Fig. 3 (a) is taken of somewhere in Shenzhen, China. We notice that the segmentation of water area is homogenous and united in Fig. 3 (b). We have 8 samples of water segments. ISODATA algorithm classifies those 8 samples into 2 different types. The final extraction is attained when ratio is set at 1.2 as Fig. 3 (c) shows.

(a)

(b)

(c)

Figure 3. Extraction Result of Trial 2: (a) is the original aerophotograph of Shenzhen, China. (b) is the result of watershed segmentation. (c) is the final extraction of water by setting the ratio at 1.2

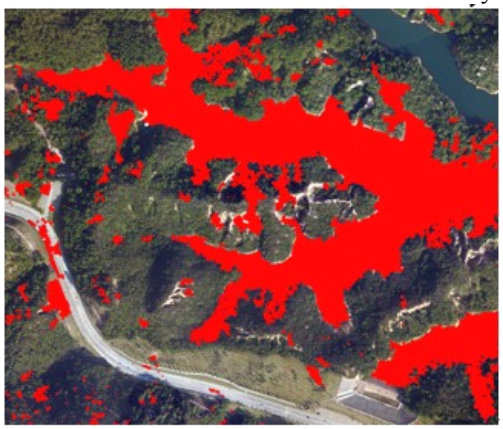
TABLE II. EVALUATION OF TRIAL 2

Fig. 3 (c)	w-w	w-n	n-w	Total Pixels
Pixels	3941135	70597	43726	4055458
Percentage	97.2%	1.7%	1.1%	100.0%

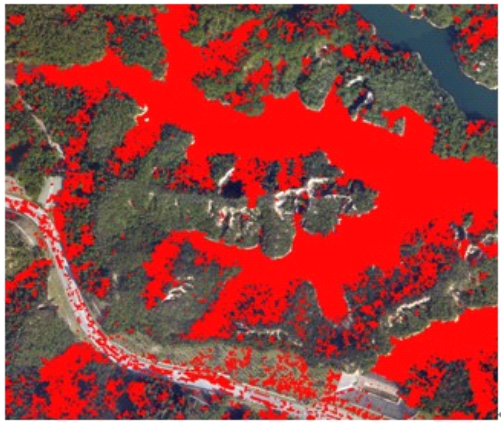
Trial 3: To test the effectiveness in the combination of chromatic and textural analysis, we conduct another extraction using the aerophotograph in Trial 1. When conducting chromatic analysis only, the feature vector will be shortened to have only a* and b* parameters. When conducting textural analysis, the feature vector will be shortened to have only ASM parameter. We also selected 8 same water areas as samples in trial 1. In chromatic extraction only, ISODATA classifies those 8 samples into 3 classes and the extraction result is Fig. 4 (a). In textural extraction, ISODATA classifies those 8 samples into 4 classes and the extraction result is Fig. 4 (b). For both trials, the ratio is set 1.2 as the ratio set in trial 1. From the result we can see that the chromatic extraction or textural extraction individually can not attain an effect as good as the combination of them. The percentage of w-n and n-w is

 ISBN: 978-1-61208-118-2

a nightmare for correctively extract water areas compared to the combination of both chromatic and textural analysis.



(a)



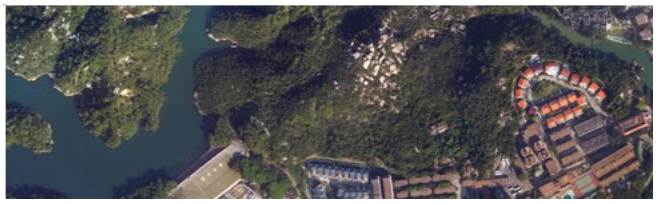
(b)

Figure 4. Extraction Result of Trial 3: (a) is attained using ony chromatic analysis when ratio is set 1.2. (b) is attained using only textural analysis when ratio is set 1.2

TABLE III. EVALUATION OF TRIAL 3

Fig. 4 (a)	w-w	w-n	n-w	Total Pixels
Pixels	2601923	391245	425432	3418600
Percentage	76.1%	11.4%	12.5%	100%
Fig. 4 (b)	w-w	w-n	n-w	Total Pixels
Pixels	2615619	398715	743726	3758060
Percentage	69.6%	10.6%	19.8%	100.0%

Trial 4: In order to test the generality of this methodology, we use the **CCV** generated in trial 1 to extract water area from another aerophotograph (Fig. 5 (a)) taken of the same area. Because they are the same region, therefore their chromatic and textural features are similar. The following result (Fig. 5 (b)) is attained when ratio is set 1.2.



(a)

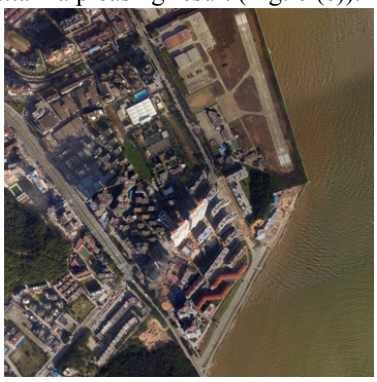
(b)

Figure 5. Extraction Result of Trial 4:(a) is the orginal RGB aerophotograph. (b) is attained using the **CCV** generated in trial 1 when ratio is set 1.2.

TABLE IV. EVALUATION OF TRIAL 4

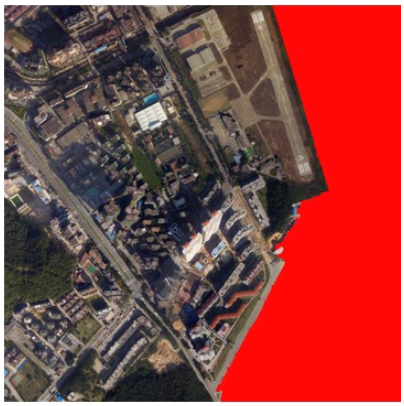
Fig. 5 (b)	w-w	w-n	n-w	Total Pixels
Pixels	1437772	26133	10216	1474121
Percentage	97.5%	1.8%	0.7%	100.0%

Trial 5: In order to further test the generality of this methodology, we apply it to extract water areas from aerophotograph with many fluctuations. Following aerophotograph (Fig. 6 (a)) is taken of somewhere in Qingdao. The biggest characteristic of this aerophotograph is its rippled surface. As a result of this irregular fluctuation, the segmentation result (Fig. 6 (b)) is not very satisfactory. There are many fragments in water area, but we can manually merge those fragments into united one in order to improve the result of segmentation. We choose 12 water samples from Fig. 6 (b). ISODATA algorithm automatically classifies those samples into 4 categories. When ratio is set 1.5, we can attain a pleasing result (Fig. 6 (c)).



(a)

(b)



(c)

Figure 6. Extraction Result of Trial 5: (a) is the orginal RGB aerophotograph. (b) is the result of segmentation. (c) is the final result of extraction.

TABLE V. EVALUATION OF TRIAL 5

Fig. 6 (c)	w-w	w-n	n-w	Total Pixels
Pixels	316823	10192	716	327731
Percentage	96.7%	3.1%	0.2%	100.0%

VII. CONCLUSION

It is innovative to combine watershed segmentation, CIELAB color space and chromatic and textural analysis as well as ISODATA classification together to extract water from RGB aerophotograph. Through numerous experimental comparisons, we discover that ratio set between 1.1 and 1.8 is optimal for water extraction. However, it is undeniable that ripples or ship transportation will cause a few fragments after segmentation, which put sand in the wheel of chromatic analysis and textural analysis because feature values of those fragments are not accurate and not representative. Although merging those fragments into united and homogenous areas is a good solution to this problem, it demands much more time and manual involvements. This is what we could do to improve the result of water extraction later on.

ACKNOWLEDGMENT

Special thanks to Professor Yongjun Zhang's guide in this research and the cooperation of our team members.

REFERENCES

[1] J. Deng, "An Effective Way for Automatically Extracting Water Body Information from SPOT-5 Images", Journal of Shanghai Jiaotong University (Agricultural Science), vol. 23, no. 2, pp. 198-201, 2005, June.

[2] H. Xu, "A Study on Information Extraction of Water Body with the Modified Normalized Difference Water Index(MNDWI)", Journal of Remote Sensing, vol. 9, no. 5, pp. 589-595, 2005, September.

[3] W. Ma and B. S. Manjunath, "A Texture Thesaurus for Browsing Large Aerial Photographs", Journal of the American Society for Information Science, CA, vol. 49, no. 7, pp. 633-648, 1999.

[4] Niblack W, Barber R et al., "The QBIC project: Querying images by content using color, texture and shape", Proc of SPIE: Storage and Retrieval for Image and Video Database, San Jose, CA, pp. 58-70, 1994.

[5] A. R. Robertson, "Historical development of CIE recommended color difference equations", Color Research and Application, 15(3), pp. 458-461, 1990.

[6] K. R. Castleman, Digital Image Processing, New Jersey: Prentice Hall, Inc., a Simon&Schuster Company, pp. 499, 1996.

[7] B. Julesz, "Experiments in the visual perception of texture", Scientific Americian, vol. 232, pp. 34-43, 1992, May.

[8] R. M. Haralick and K. S. Dinstein, "Textural Features for Image Classification", IEEE Transactions on Systems Man and Cybernetics, pp. 610-621, 1973, November.

[9] J. R. Jensen, Introductory Digital Image Processing-A Remote Sensing Perspective, 3rd ed., New Jersey: Prentice Hall, pp. 383-389, 1995.

 ISBN: 978-1-61208-118-2

O*: A Bivariate Best First Search Algorithm to Process Optimal Sequence Traversal Query in a Graph

Qifeng Lu
MacroSys LLC.
Arlington, United States
qilu1@vt.edu

Kathleen Hancock
Center for Geospatial Information Technology, Virginia Polytechnic Institute and State University
Alexandria, United States
hancockk@vt.edu

***Abstract*—An Optimal Sequence Traversal Query is a new query in a graph typically representing a transportation network that determines the minimum-cost path with a predefined origin-destination pair, traversing a set of non-ordered points of interest, at least once for each point. It has distinctive applications in the GoeProcessing domain in transportation where a user may query a shortest route to start from his/her office, traverse a set of consumer destinations, and then go home. Optimal Sequence Traversal Query generalizes TSP in the sense that the origin and the destination may be different in the former. This paper proposes a bivariate best first search, O*, to process such a query within a graph. Two special cases of O*, O*-SCDMST and O*-Dijkstra, are provided, their performance in a fully connected directed graph are studied through a set of experiments, and the result demonstrates that, on average, O*-SCDMST reduces computation time by one order of magnitude when compared to O*-Dijkstra.**

Keywords- Bivariate Best First Search, Heuristic, Optimal Sequence Traversal Query, O*, O*-SCDMST

I. INTRODUCTION AND BACKGROUND

The Optimal Sequence Traversal Query (OSTQ) is a query conducted to find the minimum-cost path that starts from any given origin, passes through a set of points of interest, and terminates at a given destination. It has distinctive applications in the GeoProcessing domain in transportation where a user may query a shortest route to start from his/her office, traverse a set of consumer destinations, and then go home. It may have potential applications in artificial intelligence where a robot is sent to collect data from multiple sensors and delivers the data to a given destination for downloading and analysis. The traveling salesman problem (TSP) is a special case of OSTQ, where the given origin and destination are the same [1] [2] [3] [4] [5] [6].

A graph can conceptually represent a complex network, such as a transportation network. In this context, a graph G is defined as a set of vertices, $\{V_i\}$, and a set of directed line segments, $\{E_{i,j}\}$, called arcs. $E_{i,j}$ is defined such that an arc is from vertex V_i to vertex V_j, and V_j is a successor of V_i. Each $E_{i,j}$ has an associated cost $C_{i,j}$. Only graphs with $C_{i,j}>=0$ are considered in this paper. These graphs are referred to as δ graphs.

In this paper, we propose a new type of query defined in such a δ graph: Optimal Sequence Traversal Query (OSTQ). Given a δ graph G with its vertices $\{V_i\}$ and edges $\{E_{ij}\}$, a set of vertices of interest VI from $\{V_i\}$ in graph G, a starting vertex S, and a destination vertex D, an OSTQ retrieves the minimum-cost path, traversing all vertices of interest, at least once for each vertex. In such a graph, there may exist *normal vertices* that are neither vertices of interest to traverse, nor the given origin or destination. Within this context, all necessary sub state graphs are generated implicitly. An application of this query in the GeoProcessing domain is that a consumer drives from his/her office, traverses a gas station, a coffee shop, and a post office, and gets home. Another application is that a food delivery person starts from the restaurant he works at, traverses a set of delivery points, and then returns to the restaurant, a typical example of TSP.

This paper proposes a bivariate best-first search algorithm, O*, to process OSTQ in a graph. O* is a bivariate best first search in the sense that it uses two variables to specify its state. Both theoretical and experimental analyses of the proposed algorithm are presented.

The paper is organized as follows. First, related work is discussed in Section II. In Section III, O* is presented. Section IV provides the SCDMST heuristic, a globally admissible heuristic for O*. Section V presents a set of experiments and an analysis of the results. Finally, the conclusion is presented.

II. RELATED WORK

This section provides a review of the state-of-the-art research on 1) Traveling Salesman's Problem (TSP), and 2) best first searches.

A. Travelling Salesman Problem (TSP)

The earliest research in TSP is in Euclidean space that searches for a shortest round-trip route to traverse each city exactly once with all cities directly connected to each other, forming a fully connected graph. A set of solutions, including dynamic programming [7], nearest neighbor [8], iterative algorithms such as 2-OPT, 3-OPT, and n-OPT [9], best first search [10], ant colony simulation [11], simulated annealing [12], Branch and Bound approach [13], and so on, were proposed to resolve this problem, either exactly or approximately, and the result is a Hamiltonian cycle that visits each vertex exactly once and returns to the starting vertex. These algorithms may be adjusted to process OSTQ.

B. Best First Searches

A best first search is a kind of informed search. The following two subsections provide a review of best first searches. A best first search is n-variate if it uses n variables to specify its states.

 ISBN: 978-1-61208-118-2

Single-Variate Best First Searches

Single-variate best first search is the existing best first search that searches a graph by expanding the most promising vertex chosen according to some rule. It adopts estimates to the promise of vertex n by a "heuristic evaluation function *f(n)* that, in general, may depend on the description of *n*, the description of the goal, the information gathered by the search up to that point, and most important, on any extra knowledge about the problem domain." [14], which is in prevalent used by researchers in Artificial Intelligence (AI), including Russell & Norvig [15].

Several algorithms, including A* [15][16], Dijkstra search [17], Greedy search [15], frontier search [18], and so on, extract the path of minimum cost between a predefined origin-destination vertex pair in a graph. A* uses a distance-plus-cost heuristic function as *f(n)* to determine the order, in which the search visits vertices in the graph [14]. *f(n)* is the sum of two functions: *g(n)*, the path cost function of the path from the origin to the current vertex *n*, and *h(n)*, the heuristic estimate of the distance from the current vertex *n* to the goal. For *h(n)*, two important concepts exist. The first is *admissibility*. A heuristic is *admissible* if its value is less than or equal to the actual cost [15]. The other is *consistency*. A heuristic is *consistent* when the real cost of the path from any vertex *A* to any vertex *B* is always larger than or equal to the reduction in heuristic [16]. Once a heuristic is consistent, it is always admissible [15]. A* has been shown to obtain the optimal solution, i.e., the minimum-cost solution, whenever the heuristic is admissible [15], and, indeed, is optimally efficient among all best-first algorithms guided by path-dependent evaluation functions when the heuristic is consistent [19]. Additionally, A* is complete in the sense that it will always find a solution if one exists. The spatial and time complexity of A* depends on the heuristic and, in general, is exponential. However, A* is very fast in practice. Both Dijkstra and the Greedy algorithm can be considered as a special case of A*. Dijkstra algorithm only considers g(n) as f(n). The Greedy algorithm only uses h(n) as f(n). Frontier search is similar to A* except that Frontier search only works on data sets with consistent heuristic and does not require a closed-list to implement the search algorithm, which consequently saves space at the cost of increased computation [18].

There is also a set of A* variations such as anytime A* [20], hierarchical A* [21], MA* [22], and SMA* [23], which take the same form *f(n)* as A* but adapts A* to different scenarios to reduce time or space complexity of A*.

All the *f(n)*s used in these identified best first searches are defined upon a single variable, vertex n, to estimate its promise.

One exact approach uses a Minimum Spanning Tree (MST) to provide an admissible heuristic to retrieve optimal TSP routes with A* [10]. The algorithm's performance has not been reported since then. A possible reason is that to process TSP, existing single-variate best first search is not adequate. This is because a vertex must be able to store multiple partial paths that traverse different sets of points of interest during the search process that a single-variate best first search cannot handle.

Bivariate Best First Searches

The concept of multivariate best first searches was first proposed in [24] to address the deficiency of a single-variate best first search to process multiple categories of interest. It uses multiple variables to specify a state to be evaluated and expanded. L#, a generalized best first search that evaluates the promise upon a state in a similar form as A*, was proposed, together with a set of novel concepts in best first searches, including local heuristic, global heuristic, local admissibility, and global admissibility [24]. As an instance of L#, the bivariate best-first-search C* was provided to processes Category Sequence Traversal Query (CSTQ) in a graph, which asks for a minimum cost route that starts from a given origin, traverses a set of ordered categories of interest that includes multiple objects in each category, with one selection from each category, and ends at a given destination [24]. In C*, a bivariate state instead of a single-variate state is evaluated and expanded. The state in C* is defined as the combination of a vertex and its *VisitOrder*, a discrete integer variable to indicate the order of a visiting category. A vertex may have multiple states in a graph, and not all the states of a vertex may be generated and expanded. Through its state specification, C* extends the theorems on optimality identified in single-variate A*.

As a bivariate best first search, C* substantially improves the ability of best first searches to process more complex queries. However, C*'s state specification is still not adequate to process OSTQ because it does not consider different visit orders of a vertex of interest. Therefore, C* cannot be used to process OSTQ.

Since best first search is supported with solid theories to obtain optimally efficient, optimal, and sub-optimal solutions and in nature a Branch and Bound approach that is prevalent for TSP calculation, a bivariate best first search is proposed in this paper to process OSTQ.

III. O*: A BIVARIATE BEST-FIRST-SEARCH APPROACH TO PROCESS OSTQ IN A GRAPH

This section describes the details of O*, the bivariate best first search algorithm to process OSTQs in a graph. First, for a state *s*, O* uses the same-form distance-plus-cost function *f(s)* as C*, shown in equation (1), to determine the order, in which the search visits vertices in the graph [24]. Therefore, similar to C*, O* is another instance of L#. Second, O* assures that its solution traverses all points of interest.

$$f(s)=g(s)+h(s) \quad (1)$$

where

f(s) is the estimate to the promise of a state *s* to be expanded,

g(s) is the cost from the origin state to the state *s*, and

h(s) is the estimation to the actual cost from the state *s* to the final state.

The lower the *f(s)*, the higher is the priority for a state to be expanded.

In the following subsections, first, a set of best first search concepts identified for O* is discussed, followed by the presentation of the O* algorithm. Next, how the algorithm assures the traversal of all vertices of interest is

 ISBN: 978-1-61208-118-2

discussed, followed by the analysis of its completeness. Finally, the relationship between different states of a vertex is discussed.

A. Definition

In O*, a state $S_{i,j}$ is defined as (V_i,VL_j), where VL_j is an ordered list of j vertices of interest that are traversed along the path obtained so far at vertex V_i. Since a vertex may have multiple states and each state has its own points of interest traversed, the ordered list is used to efficiently compare two states to remove a state that is not on the target route to be retrieved. A state of a vertex n describes the traversed points of interest along the partial path from the origin to the vertex n. A successor operator Γ is defined on $\{S_{i,j}\}$. Its value for each $S_{i,j}$ is a set of child states of $S_{i,j}$. Whenever a $S_{i,j}$ is of a vertex of interest, a Ψ operator will transform the state $S_{i,j}$ to $S_{i,j+1}$, (V_i,VL_{j+1}), at no cost by adding the vertex of interest to VL_j. A state graph *SG* defined on a graph *G* is the graph *G* whose vertices are of the same *VL*. The number of state graphs for *G* is the number of *VL*s. Since a *VL* is a sorted list storing traversed vertices of interest, in a problem with n vertices of interest, the number of state graphs is 2^n. These state graphs compose the state graph space $\mathcal{G}$, an item-enumeration graph space. A sub state graph *SSG* is a portion of a state graph *SG,* and a forest in nature. It is said to be *implicitly* generated by Γ operations in *SG*. A sub state graph space G_s is a set of sub state graphs, i.e., *SSG*s. In G_s, each point represents a sub state graph that contains the expanded vertices to each of which the path from the origin has traversed the same set of vertices of interest described by an item enumeration variable. It is said to be *implicitly* generated if it is initiated with a single source state $S_{0,0}$ and a set of Γ operations and some possible Ψ operations applied to it, to its successors, and so forth. A δ sub space graph space is a special G_S with edge cost always not smaller than 0. A path from a source *S* to a goal *D*, traversing a set of vertices of interest, is an ordered set of states (V_i,VL_j). In O*, all sub state graphs are generated implicitly. The concepts are illustrated by the example in Section IV.*B*.

Since OSTQ traverses multiple vertices of interest between the origin and the destination, the way to estimate its heuristic is similar to in C* but different from that in A* that is directly based on the currently generated vertex and the destination. Since O* is an instance of L#, the following concepts identified for L# [24] are also applicable to O*: *local heuristic, global heuristic, local admissibility,* and *global admissibility.* For these concepts, the only difference between in C* and in O* is the state used in their corresponding definitions. For example, in O*, a global heuristic, *hg*, is estimated based on the currently generated vertex, the remaining vertices of interest to traverse, and the final goal, while C*'s global heuristic is estimated based on the remaining categories of interest to traverse instead of the remaining vertices of interest to traverse [24]. The following describes these concepts in O*.

Local heuristic is the estimate to the actual cost from the current vertex to a *subgoal, sg*. A subgoal is a vertex that is a point of interest (PoI). For two vertices n and n', *local consistency* means the following inequality exists:

$$hl(n,sg) <= g^*(n, n') + hl(n', sg) \quad (2)$$

where
hl is a local heuristic, and
$g^*(n, ,n')$ is the actual cost from *n* to *n'* in the graph.

Global heuristic is the estimate to the actual cost from the current state to estimate to the final goal.

Global admissibility means the global heuristic is not larger than the actual cost from the current state to evaluate to the final goal state.

B. The O Algorithm*

O* incrementally searches all paths leading from the starting vertex, traversing the vertices of interest, until it finds a path of minimum cost to the goal. It first takes the paths most likely to lead towards the goal.

Similar to C*, O* also maintains a set of partial solutions, unexpanded leaf states of expanded vertices. These solutions are stored in an *open list*, also called a *priority queue*, which is a sorted queue based on each element's priority. Same as in C*, the priority is assigned to a state *s* based on the function (1).

Even though C* and O* use the same-form bivariate distance-plus-heuristic function *f(s)*, they use different state definitions.

The lower the *f(s)*, the higher is the priority for a vertex to be expanded. Whenever an equal f(s) occurs, the state with a larger *VisitList* will be the next to expand. Otherwise, one is randomly selected. State *A*'s *VisitList,* vl_A, is *larger* than State *B*'s *VisitList*, vl_B, i.e., $vl_A > vl_B$, if the length of vl_A is longer. In other words, the path from the start state to *A* traverses more vertices of interest than that to *B*.

For an N-point traversal problem, O* first generates the source state that contains the given vertex and an empty *VisitList*, and puts it into the open list. For all the states in the open list, the algorithm expands the state with the lowest *f(s)* value, and its children states are generated. A child state always inherits the *VisitList* of its parent whenever the child is not a vertex of interest; otherwise, the child's *VisitList* will be incremented by adding the vertex to it. The process continues until a goal state whose vertex is the final goal and *VisitList* contains all the vertices of interest or no solution is found. Once a goal state is reached, it will retrieve the obtained path using a data structure called *backpointer*, the combination of vertex identification and *VisitList,* to recursively obtain the parent until the origin state is reached.

1) O Pseudo Code*

Given an estimation function for *hg(s)*, the starting vertex *S*, the goal *D*, and the vertices of interest to visit, the pseudo code for O* is provided in Figure 1.

2) Time and Space Complexity

In O*, the complexity between a vertex and a subgoal is the same as in A*, whose time complexity and space complexity are dependent on the heuristic. For A*, in general, the time and space complexities are both exponential. Assume *b* is the branching factor, *d* is the largest depth to obtain the shortest path, and then both the time complexity and the space complexity are $O(b^d)$ [15]. A* is sub-exponential only when its heuristic *h(x)* and the actual cost *h*(x)* satisfies the following condition [15]:

```
Input:
Starting vertex S, Goal vertex D, VisitList: a sorted list to store the traversed vertices of interest
bAdd2PQ=false: Indicator that indicates whether a generated vertex is put into PQ, not be put into PQ by default
Priority queue PQ(=set of generated (s(vertex,VisitList), f(s),g(s))) begins empty.
Closed list CL (= set of previously visited (s(vertex,VisitList), backpointer, f(s),g(s))) begins empty.
Algorithm Process:
Put (S, VisitList =null), f=hg(S,null), and g(S,null)=0 into PQ
While (PQ is not empty)
{     remove the state with the lowest f having the largest VisitList from PQ. Name it n.
    If n is a goal, then  //a goal must be the predefined destination after all  vertices of interest are visited
    {       Succeeded, report the result, and END;
    }
    Else
    {       put n with its f,g,and backpointer in CL
          For each v' in successors(n.vertex)
           {      if v' is an unvisited sub goal, then
                        VisitList'=(n.VisitList).add(v'); // add n to the VisitList
              Else
              {         VisitList'=n.VisitList;
                        g(v',VisitList')= g(n)+Cost(n.vertex,v');
                        hg (v',VisitList')=calculateGlobalHeuristic(v',VisitList');
                        f (v',VisitList')=g(v',VisitList')+hg(v',VisitList')
                        Process((v',VisitList'), f, g) //decide to place the state in PQ and/or to remove other states from CL/PQ
              }
           }
    }
}
Process((v',VisitList'), f, g):
   bAdd2PQ=true;
    If v' not seen before, or (v',VisitList') currently in PQ with f(v',VisitList')>f Then
          Place/promote (v', VisitList') on priority queue with f ,g; END;
    If (v', VisitList*) is in PQ, Then
          If(VisitList* is a super set of  VisitList' && g(v',VisitList*)<g(v',VisitList')) Then
                bAdd2PQ=false;
          If(VisitList* is a sub set of VisitList' && g(v',VisitList*)>g(v',VisitList')) Then
                Delete (v',VisitList*) from PQ
    If(v',VisitList') previously expanded Then
          If f(v',VisitList')<=f Then
                bAdd2PQ=false;
          else
                Delete (v',VisitList*) from the closed list
    Else
        If(v',VisitList*) is in CL Then
                If(VisitList* is a super set of  VisitList' && g(v',VisitList*)<g(v',VisitList')) Then
                      bAdd2PQ=false;
                If(VisitList* is a sub set of VisitList' && g(v',VisitList*)>g(v',VisitList')) Then
                      Delete (v',VisitList*) from CL
   If(bAdd2PQ) Then Place (v',VisitList') with f,  and g on priority queue
```

Figure 1. The pseudo code of O*

$$|h(x)-h^*(x)|<=O(\log h^*(x)) \quad (3)$$

Where

h(x): the heuristic in A*, i.e., the local heuristic in O*; and

h(x)*: the corresponding actual cost

For O*, assume b is the branching factor, d is the largest depth to obtain the shortest path between two objects including the origin, the destination, and the specified vertices of interest to traverse, which is named as a *section path* in O*, m is the number of vertices of interest specified to traverse, and F_h is the worst time complexity to obtain a global heuristic, then the state length is $m+1$, the number of state graphs is 2^m. In the worst case, O* requires traverse paths resulting from all possible order combinations of vertices of interest and storage of all state graphs. Each section path is exponential in time and space complexity, and the corresponding time complexity is $O(m!F_h b^d)$, and space complexity is $O(m2^m b^d)$.

To reduce the time complexity to sub-exponential, the heuristics used in O* must be sufficiently close to the actual cost to reduce the number of candidate order combinations from *of permutation* level to *of sub-exponential* level.

3) Traversal Constraints of Vertices of Interest

The solution from O* must traverse all the vertices of interest to be a candidate solution to an OSTQ.

Lemma 1: The solution from O* satisfies the traversal constraint of vertices of interest.

Proof:

Use contradiction.

Assume the solution obtained from O* does not traverse at least one vertex of interest.

Since the solution is obtained, then the search stops. According to O*, if a subgoal, a specified vertex of interest, is not reached, then it cannot be added to the *VisitList* of any

vertex generated by O*, and thus there is no state whose *VisitList* contains the vertex of interest. Consequently, the final state will never be reached. In other words, the search will not stop to report a solution if there is one. This is contradicted with the assumption. So O* does provide a solution that satisfies the constraint to traverse the specified vertices of interest.

C. Completeness

Completeness means that an algorithm finds a solution if one exists.

Theorem 1: O* is complete.

The algorithm will not stop until either the goal is reached or there is no solution to the OSTQ.

D. Admissibility and Optimality

Admissibility is important in A* since it guarantees the solution is optimal. This is also true in O*.

Lemma 2: If any global heuristic for any vertex is admissible, then the solution is optimal.

Proof:

Use contradiction.

Suppose O* finds a suboptimal path, ending in goal state (G_1,vl_g) where vl_g contains all the vertices of interest since the search guarantees the solution traverses all the vertices of interest, i.e., $f(G_1,vl_g) > f^*$ where $f^* = hg^*$ *(origin state)* = cost of the optimal path. Let $hg^*(n,vl)$ as the actual cost from a state (n,vl) (n is the vertex and vl is its *VisitList*) to the goal state.

There must exist a state (n,vl) that is unexpanded, to which the path from the origin state is the start of a true optimal path, and $f(n,vl) >= f(G_1, vl_g)$ (else search would not have ended).

Also $f(n,vl) = g(n,vl) + hg(n,vl) = g^*(n,vl) + hg(n,vl)$ because (n,vl) is on the optimal path,

Since the global heuristic is globally admissible, then

$f(n,vl = g^*(n,vl) + hg(n,vl)) <= g^*(n,vl) + hg^*(n,vl) = f^*$

So $f^* >= f(n,vl) >= f(G_1, vl_g)$

Contradicting the assumption. So the solution is optimal.

Once the global heuristic is admissible, then the solution is optimal.

An algorithm is defined as *admissible* if it is guaranteed to find an optimal path from s to the goal state for any δ graph, traversing at least once for each specified vertex of interest.

Theorem 2: Once any global heuristic is admissible, then O* provides the optimal solution to the corresponding OSTQ, i.e., O* is admissible.

Proof:

First, based on Lemma 1, O* guarantees that each specified vertex of interest is traversed at least once. Then based on Lemma 2, global admissibility guarantees solution optimality. Consequently, an optimal solution that satisfies the vertices of interest traversal constraint is the optimal solution to the corresponding OSTQ, completing the proof.

E. Relationship between Different States of a Vertex

Since in a bivariate best first search, a vertex may have multiple states, the relationship between any two of its states can be used to prune unnecessary states.

Theorem 3: For a vertex v,

If(g(v,VL')>g(v,VL) AND VL is a super or equal set of VL', then the state (v,VL') is not on the optimal path.

Proof:

Since *VL* is a super or equal set of *VL'*, which means *VL* contains at least the same set of traversed vertices of interest as *VL'*, it is clear that $g^*(v,VL)>=g^*(v,VL')$. According to the given condition, $g(v,VL')>g(v,VL)$, then $g(v,VL')>g^*(v,VL)$, so $g(v,VL')>g^*(v,VL')$. Consequently, (v,VL') is not on the optimal path. Completing the proof.

IV. SCDMST: TO PROVIDING A GLOBALLY ADMISSIBLE HEURISTIC TO PROCESS OSTQ IN A FULLY CONNECTED DIRECTED GRAPH

According to the time complexity analysis in Section 3.2.2, it is desirable to reduce the depth to process OSTQ to reduce the computation time in the worst case, especially for a program within a graph where the majority of the vertices are not points of interest. Through Dijkstra, an algorithm using polynomial time in terms of all the number of vertices in a graph to obtain routes for a single-source multiple-destination routing problem, it is clear that an OSTQ processing in a general graph that contains both vertices of interest and general vertices can be efficiently transformed into a fully connected graph containing only the points of interest plus the origin and the destination. This is also the reason that TSP is primarily studied in a fully connected graph. In addition, directed graphs are more general in real world trip planning. Therefore, this section presents an instance of O* that uses a global heuristic to process OSTQ processing in a fully-connected directed graph.

Consider a directed graph, $G(V,E)$, where V and E are the set of vertices and edges, respectively, and a starting vertex s and an ending vertex e in E. Associated with each edge (i,j) in V is a cost $c(i,j)$. Let $|V|=n$ and $|E|=m$. A semi-connected directed spanning tree, *SCDST*, is defined as a graph that connects, without any cycle, all vertices with $n-1$ arcs, while vertex s only has outgoing arcs and vertex e only has incoming arcs. It is semi-connected in the sense that it does not necessarily connect any two points together. A Semi-Connected Directed Minimum Spanning Tree (*SCDMST*) is the graph with the minimum total edge cost among all *SCDST*s. In other words, the problem is to find a *SCDST*, $G(V,S)$ where S is a subset of E, such that the sum of $c(i,j)$ for all (i,j) in S is minimized. Figure 2 shows a *SCDMST* example.

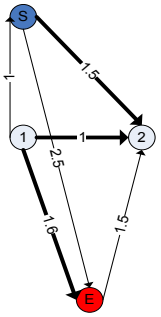
The *SCDMST* heuristic is globally admissible in a fully connected directed graph whose edge costs obey the triangle inequality. The property is proved through Theorem 4.

Theorem 4: A *SCDMST* heuristic is globally admissible in a fully connected directed graph whose edge costs obey the triangle inequality.

Proof:

Use contradiction.

Suppose the *SCDMST* heuristic is larger than the actual cost. Therefore, in the *SCDMST* there must exist at least one cost between two points is larger than their actual cost. At the same time, since edges obey the triangle equality and an



Where

S: starting vertex, E: ending vertex.
The *SCDMST* tree is highlighted in dark black. The *SCDMST* starts from S, connects vertex 1 and vertex 2, and ends at E. No cycle exists. 3 edges are used to connect 4 vertices. Only outgoing edges exist for S and incoming edges exist for E. The *SCDMST* is semi-connected since not any two points are connected through the tree. For example, 2 and E are not connected together in the obtained *SCDMST*.

Figure 2. A SCDMST example

edge on the optimal path cannot be a directed edge with its ending vertex as the starting vertex of the *SCDMST* tree or with its starting vertex as the ending vertex of the *SCDMST* tree, this means an edge with a less cost is not found by the *SCDMST*, which is contradicted with the fact that a *SCDMST* always adopts the edge of the minimum cost between two points. Completing the proof.

The O* algorithm that uses a *SCDMST* to provide a global heuristic is named as *O*-SCDMST*.

A. The D-ODPrim Algorithm to Retrieve a SCDMST from a Fully Connected Directed Graph

In this paper, D-ODPrim is proposed to obtain a *SCDMST* from a directed graph. Its pseudo code is provided in Figure 3.

Input: A connected directed weighted graph with vertices V and edges E, the starting vertex s, and the ending vertex e.
Initialize: V_{new} = {x}, where x is the starting vertex from V, E_{new} = {}
Remove incoming edges of s and outgoing edges of e in E
Repeat until V_{new} = V:
  Choose edge (u,v) from E with the minimal weight such that one vertex is in V_{new} and the other is not
  (if multiple edges with the same weight exist, choose arbitrarily)
  Add v to V_{new}, add (u, v) to E_{new}
Output: V_{new} and E_{new} describe a semi-connected directed minimal spanning tree

Figure 3. The pseudo code for D-ODPrim algorithm

D-ODPrim can be regarded as a variation of the Prim algorithm that calculates a MST for an undirected graph [25]. Similar to Prim's algorithm, D-ODPrim continuously increases the size of a tree starting with the given starting vertex until it spans all the vertices.

Next, we prove the algorithm outputs a SCDMST.

Proof:

Four steps are used to prove D-ODPrim outputs a SCDMST if a solution exists.

Assume the number of vertices is *N*. Name the obtained graph as *DG*.

Step 1: Prove that *DG* contains *N*-1 edges.

Proof: Every time a new edge connecting to a new vertex is added to E_{new}, until the *N*-1 points are added to V_{new}. Consequently, for N vertices, the algorithm will try *N*-1 times, thus *N*-1 edges will be added to form *DG*.

Step 2: Prove that *DG* does not contain a cycle.

Proof: According to the algorithm, if the edge direction is neglected in *DG*, *DG* connects any two points together with *N*-1 edges, which is impossible to have a cycle. So the directional *DG* does not contain a cycle.

Step 3: Prove that no outgoing edges for the ending vertex e and no incoming edges for the starting vertex s.

Proof: Since the algorithm removes the incoming edges of s and the outgoing edges of e from the candidate edge set, it is no way to add these edges into *DG*.

Step 4: Prove that the total edge cost is the minimum.

Proof: Use contradiction. Suppose another *SCDST*, DG_1, has a smaller total edge cost. Then there must be at least one vertex that uses an edge of smaller cost in DG_1 than in *DG*, which conflicts the fact that every time the algorithm finds the minimum cost edge for a vertex to add it to *DG*.

Based on the four steps, it is clear that D-ODPrim retrieves a SCDMST, completing the proof.

For D-ODPrim, the time complexity is $O(V^2)$ and space complexity is $O(V^2)$.

B. An Example

To illustrate how O*-SCDMST works, the following simple example is provided. Figure 4 shows the fully connected directed graph having C_{ij}=C_{ji} where i and j represents any two vertices in the graph and C_{ij} represents the cost from i to j. The OSTQ ask for a shortest route starting from *O*, traversing vertex *1* and vertex *2*, and then ending at *D*. The edge costs in the graph obey triangle inequality. Therefore, O*-SCDMST retrieves an optimal solution for the OSTQ problem, and its search process in also shown in Figure 4.

The algorithm begins with the starting point *O* with *VisitList* as null and parent (*parentpointer)* as null. Its *f* value is 3.5, the cost of SCDMST of *O*,*1*,*2*, and *D*. The state is put into the priority queue. Then, the state of the lowest *f* value, vertex *O* with *VisitList* null is expanded first and put into the closed list with a null *backpointer*, and its three children, *1*, *2,* and *D*, are generated. *D* is not a subgoal, so its *VisitList* is null, inheriting its parent. Both *1* and *2* are subgoals, so their *VisitList*s are changed accordingly. For each of these three vertices, O* calculates its heuristic based on its corresponding SCDMST, and then calculates its *f* value. All generated states are put into the priority queue.

Next, since state *(1,(1))* has the lowest *f*, 3.5, it is expanded and put into the closed list with a backpointer pointing to *(O,null)*, and its children *O,2,and D* are generated. Neither *O* nor *D* is a subgoal, so their *VisitList*s are *(1)*. *2* is a subgoal, so its *VisitList* is changed to *(1,2)*. The generated three states are put into the open list.

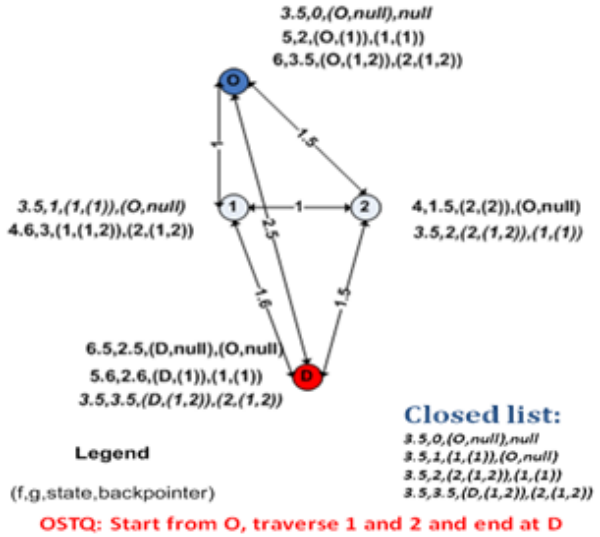


Figure 4. The OSTQ problem and O*-SCDMST search process to retrieve an optimal solution

Now *(2,(1,2))* has the lowest *f*, but it does not reach the goal state yet. So *(2,(1,2))* is expanded. Its children, *O,1,* and *D,* are generated. Now *D* is the final goal since the partial path traverses *1* and *2*. Neither *O* nor *1* is a subgoal in this case. Again, these generated states are put into the open list.

Again, based on the lowest *f*, *(D,(1,2))* is the next to be expanded. Since it is a final goal, the search stops and reports the optimal solution is O->1->2->D with 3.5 as the minimum cost.

In Figure 4, the elements in normal style are in the open list; those in italic style were in the open list first, and then moved to the closed list. The closed list in Figure 4 is its snapshot after the final goal state is expanded by O*-SCDMST. Through this example, it is clear that O*-SCDMST must store multiple states for a vertex, which a single-variate best first search cannot handle.

Figure 5 shows the corresponding sub-state-graph space representing the search space that O*-SCDMST explores to retrieve the optimal route for the OSTQ . The sub-state-graph space is composed of 2^n sub state graphs where n is the number of PoI and n=2. Note that in this example the sub-state graph with point of interest (PoI) 2 traversed has no states. The green arrow represents the Ψ operation that transforms the search space from one sub state graph to another without any cost. Each sub-state-graph is a sub graph of the original graph, with a certain set of traversed points of interest. Within each sub-state-graph, O* starts the search with a point of interest instead of the origin, O. Each PoI is traversed following a best first way. Vertices, including *O, 1, 2,* and *D,* have multiple states. These scenarios presented in O* are clearly different from the traditional single-variate best first search.

O* was implemented with C#. The experiments were performed on a Toshiba Satellite A215 Laptop with 2.0GB memory (RAM), AMD Turion™ 64*2 Mobile Technology TL-56 1.80HZ processors, and Windows Vista™ Home Premium operating system.

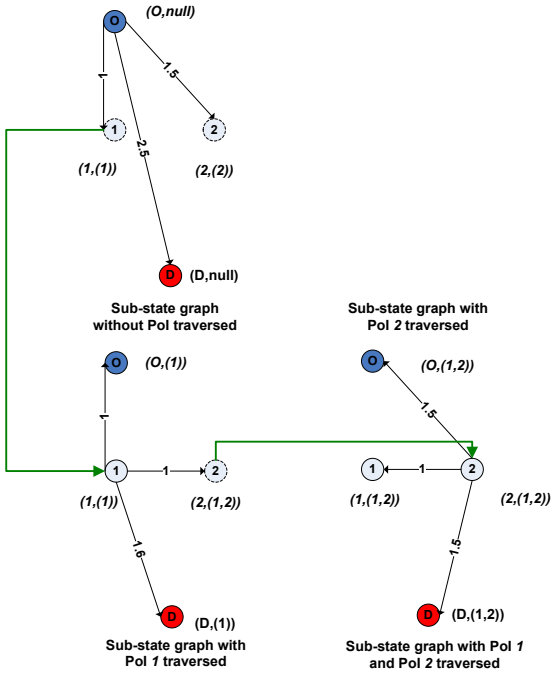


Figure 5: The sub-state-graph Space

V. EXPERIMENTS

The purpose of the experiments is to test the performance of O*-SCDMST to calculate the global heuristics in a FDG. O*-Dijkstra, a special case of O* when no heuristic is adopted, is used as the baseline. O*-Dijkstra is a consecutive network expansion algorithm to traverse multiple points of interest.

A. Data Set

An asymmetric TSP problem (Fischetti) with 34 points of interest [26], corresponding to vertices of interest in O*, is used as the data set for this experiment. The problem is a special case of Vehicle Routing Problem, and thus an asymmetric TSP [26]. The data set contains the edge costs between any two points. In this experiment, a set of OSTQ problems is generated from this data set. First, the number of points of interest consecutively changes from 2 to 15. Second, for each number of points of interest, 30 problem samples are randomly generated, i.e., the origin, the destination, and the points of interest in each problem sample are randomly selected from the 34 points. Consequently, a set of 420 problems is generated. The whole problem set is used for O*-SCDMST, the partial set with up to 12 points of interest is used for O*-Dijkstra.

B. Performance Measures

To study the performances of the two algorithms, the following performance measures are identified.

Minimum Process Time (MinPT): the minimum time required to obtain a solution for each number of points of interest (seconds);

Maximum Process Time (MaxPT): the maximum time required to obtain a solution for each number of points of interest (seconds);

Average Process Time (APT): the average time required to process a query over all runs (seconds).

C. Results and Discussion

The results are presented in Table 1. *O*-S* represents O*-SCDMST, O*-D represents O*-Dijkstra, and *NPoI* represents the number of points of interest, i.e., the number of cities to traverse. The "-" indicates that a value is not available since the time to obtain a result exceeded a reasonable expected solution time.

Figure 6 through Figure 8 are provided to visualize the performance measures provided in Table 1.

Based on MinPT, O*-SCDMST can retrieve the optimal solution within 4 seconds for an OSTQ of 15 NPoI. However, based on MaxPT, it may still require 20,858 seconds for another query with the same number of points of interest. This implies that O*-SCDMST's performance depends on how closely the selected SCDMST heuristic approaches to the actual cost.

Based on MinPT shown in Figure 6, O*-SCDMST outperforms O*-Dijkstra over all runs.

Based on MaxPT shown in Figure 7, O*-SCDMST outperforms O*-Dijkstra when NPoI larger than 2. This is due to the fact that O*-SCDMST requires additional time to compute the SCDMST heuristic, and when NPoI becomes larger, this additional time is no longer a dominant factor, instead, the obtained heuristic expedites the overall search process.

TABLE 1: PERFORMANCE RESULTS FOR O*-SCDMST, AND O*-DIJKSTRA (IN SECONDS)

	MinPT		***MaxPT***		***APT***	
NPoI	**O*-S**	**O*-D**	**O*-S**	**O*-D**	**O*-S**	**O*-D**
2	0.00	0.00	0.04	0.02	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.01	0.01	0.00	0.01
5	0.00	0.00	0.03	0.06	0.01	0.03
6	0.00	0.02	0.15	0.28	0.04	0.14
7	0.00	0.09	0.77	1.51	0.17	0.59
8	0.05	0.23	2.21	7.31	0.53	2.29
9	0.05	0.68	7.62	45.61	1.78	9.26
10	0.04	3.12	51.96	215.84	5.63	41.11
11	0.18	4.29	314.78	786.55	29.03	137.43
12	0.39	34.10	234.94	1479.48	66.82	356.69
13	3.54	-	1109.16	-	245.07	-
14	2.06	-	3900.67	-	548.08	-
15	3.37	-	20857.75	-	2204.14	-

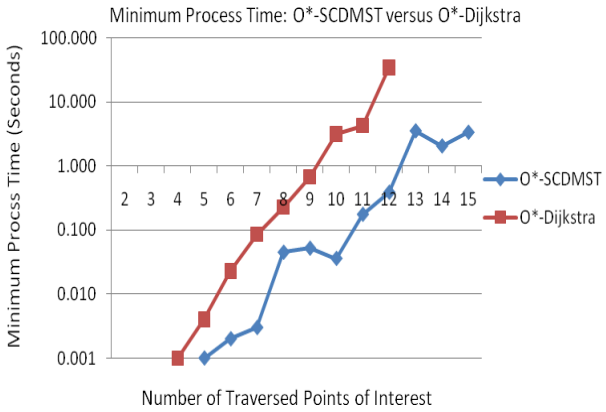


Figure 6: Minimum process time over different number of traversed points of interest: O*-SCDMST versus O*-Dijkstra

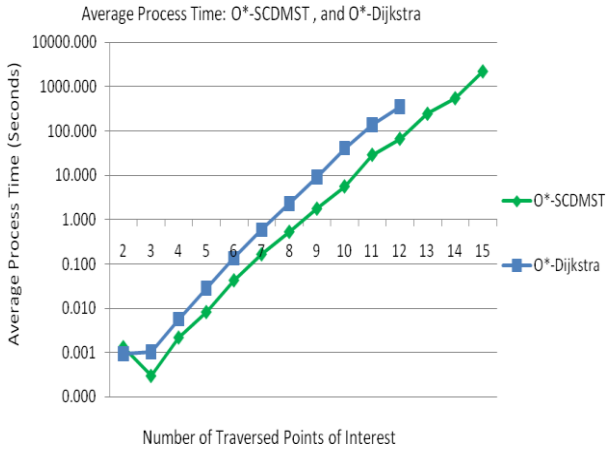


Figure 7: Maximum process time over different number of traversed points of interest: O*-SCDMST versus O*-Dijkstra

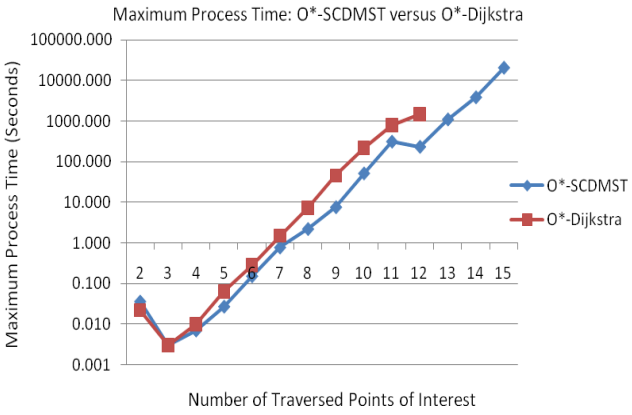


Figure 8: Average process time over different number of traversed points of interest: O*-SCDMST versus O*-Dijkstra

Based on APT shown in Figure 8, O*-Dijkstra outperforms O*-Dijkstra by an order of magnitude. On Average, O*-SCDMST can process OSTQ of up to 14 NPoI within 10 minutes.

 ISBN: 978-1-61208-118-2

In Figure 6 through Figure 8, it is noticeable that both O*-SCDMST and O*-Dijkstra are sub-exponential in time complexity.

VI. CONCLUSION

The contribution of this paper includes four components: 1) this paper proposes a novel query in a graph, OSTQ, which determines the minimum-cost path with a predefined origin-destination pair, traversing a set of vertices of interest, and provides O*, an instance of L#, to process the query in a heuristic way without explicit consideration of the order permutation; 2) A SCDMST heuristic is developed to calculate the global heuristics in O* to process OSTQ in a fully connected directed graph; 3) D-ODPrim is provided to retrieve a SCDMST for a directed connected graph; 4) The SCDMST heuristic is proved to be globally admissible if the edge costs of a fully connected directed graph obey the triangle inequality; 5) the corresponding O*-SCDMST's performance is statistically studied using a real data set. Based on the result, O*-SCDMST can retrieve an optimal solution for an OSTQ of 15 points of interest within 4 seconds at best and of 14 points of interest in 10 minutes on average. O*-SCDMST is always faster than O*-Dijkstra when the number of points of interest is larger than 2. On average, O*-SCDMST reduces the computation time by one order of magnitude compared to O*-Dijkstra.

REFERENCES

[1] S. Arora. Polynomial Time Approximation Schemes for Euclidean TSP and Other Geometric Problems. In: Proceedings of 37th IEEE Symposium on Foundations of Computer Science, Burlington, 1996, pp. 2-11.

[2] S. Arora. Approximation Schemes for NP-hard Geometric Optimization Problems: A survey. Mathematical Programming, Springer, 97 (2003) pp. 43-69.

[3] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook. The Traveling Salesman Problem: A Computational Study. Springer, 2007.

[4] N. Christofides. Worst-case Analysis of a New Heuristic for the Traveling Salesman Problem. Carnegie Mellon University, Computer Science, Tech Technical report, 1976.

[5] T. Cormen, T. Leiserson, R. Rivest, and T. Stein. Introduction to Algorithms. The MIT Press, 1997.

[6] A. Dumitrescu and J. S. B. Mitchell. Approximation Algorithms for TSP with Neighborhoods in the Plane. In: Proceedings of the 12th annual ACM-SIAM Symposium on Discrete Algorithms, Washington DC, USA, 2001, pp. 38–46.

[7] M. Held and R. M. Karp. A Dynamic Programming Approach to Sequencing Problems, Journal of the Society for Industrial and Applied Mathematics 10(1) (1962): pp. 196–210.

[8] D. J. Rosenkrantz, R. E. Stearns, and P. M. Lewis II. An Analysis of Several Heuristics for the Traveling Salesman Problem. SIAM Journal on Computing. 6 (1977): pp. 563–581.

[9] J. L. Bentley. Fast Algorithms for Geometric Traveling Salesman Problems. ORSA Journal on Computing 4, (1992), pp. 387-411.

[10] M. Held, and R. M. Karp.The Traveling-Salesman Problem and Minimum Spanning Trees. Operations Research. 18 (1970), pp. 1138-1162.

[11] Ma. Dorigo. Ant Colonies for the Traveling Salesman Problem. Université Libre de Bruxelles. IEEE Transactions on Evolutionary Computation, 1(1) (1997):pp. 53–66.

[12] E.H.L. Aarts, and J. Korst. Simulated Annealing and Boltzmann Machines: A stochastic Approach to Combinatorial Optimization and Neural Computing. John Wiley & Sons, Chichester, 1989.

[13] J. Clausen and M. Perregaard, On the Best Search Strategy in Parallel Branch-and-Bound - Best-First-Search vs. Lazy Depth-First-Search, Proceedings of the Parallel Optimization Colloquium, (1996).

[14] J. Pearl. Heuristics: Intelligent Search Strategies for Computer Problem Solving. Addison-Wesley, 1984.

[15] S. Russell and P. Norvig. Artificial Intelligence: a Modern Approach (2nd edition). Prentice Hall, 2002.

[16] P. E. Hart, N. J. Nilsson, and B. Raphael. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. IEEE Transactions on Systems Science and Cybernetics SSC4 (2) (1968) pp. 100–107

[17] E. W. Dijkstra. A Note on Two Problems in Connexion with Graphs. In Numerische Mathematik, 1 (1959), S. pp. 269–271

[18] R. E. Korf, W. Zhang, I. Thayer, and H. Hohwald. Frontier Search. Journal of the Association for Computing Machinery. 52(5) (2005) pp. 715-748

[19] R. Dechter and J. Pearl. Generalized Best-first Search Strategies and the Optimality of A*. Journal of the Association for Computing Machinery. 32(3) (1985) pp. 505-536

[20] M. Likhachev, G. J. Gordon, and S. Thrun. Ara*: Anytime A* with provable bounds on sub-optimality. In. Advances in Neural Information Processing Systems 16. MIT. Press, Cambridge, MA, 2004.

[21] A. Botea, M. Muller, J. Schaeffer. Near Optimal Hierarchical Path-Finding. In Journal of Game Development, volume 1, issue 1, (2004), pp. 7-28.

[22] S. Russell. Efficient Memory-bounded Search Methods. In Proceedings of Tenth European Conference on Artificial Intelligence, pp. 1-5. Chichester, England: Wiley, 1992.

[23] R. Zhou, and E. A. Hansen. Memory-Bounded A* Graph Search. Proceedings of the Fifteenth International Florida Artificial Intelligence Research. May 2002.

[24] Q. Lu, and K. Hancock. C*: A Bivariate Best First Search to Process Category Sequence Traversal Queries in a Transportation Network. geoprocessing, pp.127-136, 2010 Second International Conference on Advanced Geographic Information Systems, Applications, and Services, 2010.

[25] M. Fischetti, P. Toth, and D. Vigo. A branch and bound algorithm for the Capacitated Vehicle Routing Problem on Directed Graphs. Operations Research, Vol 42, pp. 846-859. 1994.

[26] Robert C. Prim: Shortest connection networks and some generalizations. In: Bell System Technical Journal, 36, pp. 1389–1401,1957

 ISBN: 978-1-61208-118-2

On Pre-Processing for Least-Cost Carpooling Routing in a Transportation Network

Qifeng Lu
MacroSys LLC.
Arlington, USA
qilu1@vt.edu

Abstract— **Within a location based social network, carpooling is becoming more and more preferable among workers both living and working near each other due to the continuous increase in gasoline price and air pollution, and is more desirable when working places are far away from homes. To find the least-cost path for carpooling means to find the optimal order to traverse the homes and workplaces to pick up passengers and to drop off them. However, before the order searching is performed, in a large transportation network, it is prevalent to first obtain the least-cost path between any two locations, with one as a work place and the other as a home, to reduce computation time. In fact, for carpool, since only a few people can stay within a vehicle, it is fast to obtain the best pickup and drop-off order. Therefore, the bottleneck to obtain the least-cost route is indeed the calculation of the least-cost paths between any such two locations. The two existing dominant approaches to pre-compute these least-cost paths are Dijkstra's algorithm and A*, where Dijkstra's algorithm is used to compute single-origin multiple-destination least-cost paths while A* is used to compute the least-cost path between an origin-destination pair. In this paper, LU, a best first search algorithm and framework recently proposed to compute single-origin multiple-destination least-cost routes, is adopted in this pre-processing step to retrieve the optimal path to traverse the carpool-based pickup and drop-off locations. Its performance is compared with A* and Dijkstra's algorithm through a set of experiments in a large transportation network. The results demonstrate that LU is significantly faster than Dijkstra's algorithm and much better than A*.**

Keywords- LU, A*, Dijkstra's algorithm, Best First Search, Single-Origin Multiple-Destination, Carpool, Location Based Social Network

I. INTRODUCTION

Within a location based social network, carpooling is becoming more and more preferable among workers both living and working near each other due to the continuous increase in gasoline price and air pollution, and is more desirable when working places are far away from homes. In general, to carpool, a worker is assigned to start from him/her home to pick up the other workers from their homes and drop off them at their workplaces and to end at his/her workplace. This carpool-based least-cost route processing is similar to a Traveling Salesman Problem (TSP) that asks for a least-cost route traversing a set of non-ordered points of interest [1][2]. Compared to TSP, it has an additional constraint in the sense that a worker must be picked up at his/her home before he/she is dropped off at his/her workplace, i.e., an order is imposed on any origin-destination pair. In addition, compared to general TSPs, there are two distinct characters for carpooling routing. First, origins and destinations are likely to cluster. Second, the number of destinations for a ride is not large.

In a large transportation network, prior to obtaining the least-cost path for carpooling, which means to find the optimal order to traverse the homes and workplaces to pick up passengers and drop off them, similar to TSP, it is prevalent to obtain the least-cost path between any two locations, with one as a work place and the other as a home, to reduce computation time [1] [2]. Otherwise, a partial optimal traversal order may have to be evaluated at each vertex along a minimum-cost route between any two locations in a transportation network, which is computation-intensive. In fact, for carpool, since only a few people can stay within a vehicle, given the pre-computed least-cost routes, it takes no time for a computer to calculate the best pickup and drop-off order. The only issue is to guarantee that the pickup location of a person is always traversed before his/her drop-off location during the optimal traversal order searching. Therefore, the dominant factor of the computation cost to obtain the least-cost route is indeed the cost to calculate the least-cost paths between any such two locations, which is the major challenge to obtain the least cost route for carpooling.

A set of web sites are available to provide people the access to carpooling and even vanpooling. erideshare.com helps diverse people groups with different trip purposes organize carpooling. Vpsiinc.com provides vanpooling with more people sharing a ride than carpooling. However, till now, none of these websites provide a complete solution for carpooling or vanpooling. In other words, no site provides services to help organize carpooling or vanpooling and calculate the optimal routes for each ride.

Three fundamental and prevalent algorithms exist to process single-origin multiple-destination routes in a graph: Dijkstra's algorithm [3], Bellman-Ford algorithm [4], and LU [5]. Compared to Bellman-Ford algorithm, Dijkstra's algorithm is more efficient but only applicable to graphs with non-negative cost edges, while Bellman-Ford can be used in graphs with negative cost edges but still cannot handle cases with negative-cost cycles. Compared to LU, Dijkstra's algorithm is less efficient when the number of destinations is relatively small compared to the total number of vertices in a transportation network [5], which is exactly the case In a carpooling scenario where there are only a small number of

 ISBN: 978-1-61208-118-2

pickup and drop-off locations in a large transportation network.

The A* algorithm is a generalization of Dijkstra's algorithm for one origin and one destination routing [6]. It takes additional information from the problem domain into account to provide a lower bound on the "distance" from a generated state to the goal. It is best-first since the vertex chosen to be expanded in a graph is the one appearing to have the shortest path from the origin to the goal.

The performance of LU has been studied against Dijkstra's algorithm [5]. However, its performance against A* has not been explored. In this paper, LU is adopted to help calculate the least cost route for carpooling with clustered destinations more efficiently when compared to existing approaches, and extensive experiments and result analysis are performed to explore the performance differences between A* and LU for a set of carpooling scenarios in a large urban transportation network. The results demonstrate that LU is significantly faster than Dijkstra's algorithm and much better than A* when the number of locations is not larger than 12.

The paper is organized as follows. First, related work is presented in Section II. Next in Section III, the algorithm LU is introduced, an example is provided to illustrate how the algorithm works, and its characteristics are discussed and compared with A* and Dijkstra's algorithm. Section IV presents the experiment and result analysis, followed by conclusions in Section V. Future research is discussed in Section VI.

II. RELATED WORK

Dijkstra's algorithm is an algorithm to retrieve the shortest paths from a single source vertex to multiple destination vertices in a weighted, directed graph [3]. All weights must be nonnegative. Dijkstra's algorithm works as follows. First, Dijkstra expands the origin and generates its children, or states, and the cost from the origin to each generated state is assigned to that state. Thereafter, Dijkstra continues to expand the state with the least cost, and generate its children, assigning the corresponding cost to each generated child. This process continues until all the destinations are reached or no state can be expanded. Dijkstra's algorithm is used widely for routing in computer network, transportation network, etc. For example, for computer network routing, Dijkstra's algorithm is the prevalent working principle behind link-state routing protocols such as OSPF and IS-IS [7][8][9]. In routing assignment in transportation, Dijkstra's algorithm plays the key role [10].

Unlike Dijkstra's algorithm, the Bellman-Ford algorithm [4] can be used on graphs with negative edge weights, as long as the graph does not contain any negative cycle reachable from the source vertex. Compared to Dijkstra's algorithm in a graph with nonnegative cost edges, Bellman-Ford requires more time to retrieve the optimal solutions.

LU, a best first search algorithm, is another approach to retrieve single-origin multiple-destination least-cost routes [5]. It uses a heuristic, *h_LU*, estimated based on destinations yet-to-be-reached to expedite the search process following a best first way. In nature, LU is a framework that can adopt different heuristics to provide optimal, optimally efficient, and sub-optimal solutions [5]. As a result, the capability of best first search was first extended to process multiple-destination queries in a graph. LU is significantly faster than Dijkstra's algorithm when the number of destinations is much smaller than the total number of vertices in a transportation network and can perform worse when the number of destinations is comparable to the total number of vertices due to the additional time to compute the heuristics.

The A* algorithm is a generalization of Dijkstra's algorithm for one origin and one destination routing [6]. It takes additional information from the problem domain into account to provide a lower bound on the "distance" from a generated state to the goal. It is best-first since the vertex chosen to be expanded in a graph is the one appearing to be the closest to the goal. A* uses a distance-plus-cost heuristic function as *f(n)* to determine the order in which the search visits vertices in the graph [6]. *f(n)* is the sum of two functions: *g(n)*, the path cost function of the path from the origin to the current vertex *n*, and *h(n)*, the heuristic estimate of the distance from the current vertex *n* to the goal.

III. LU: A BEST FIRST SEARCH ALGORITHM TO RETRIEVE SINGLE-ORIGIN MULTIPLE-DESTINATION ROUTES IN A GRAPH

In this section, the details of the recently proposed algorithm LU to retrieve single-origin multiple-destination least cost routes [5] are presented. As a best first search, LU follows a vertex generation and expansion search process. It evaluates the promise, the closeness, of each generated vertex towards the destinations yet-to-be-reached. Starting from the origin, every time LU expands the most promising vertex based on some rule and generates its children, until all the destinations are expanded, or reached [5]. It takes advantage of useful information from a problem domain to expedite the search process in a graph without negative cost edges.

A. Algorithm

LU uses the following evaluation function, *f(n)*, to evaluate the promise of a vertex, or a state, n.

$$f(n)=g(n)+h\_LU(n) \quad (1)$$

where

f(n) is the estimate to the promise of a state *n* to be expanded,

g(n) is the cost from the origin state to the state *n*, and

h_LU(n) is the minimum estimate to the actual cost from the state *n* to any unclosed destination state.

h_LU(n) is evaluated through equation (2).

$$h\_LU(n)=\min(h_i(n)) \ (1<=i<=K) \quad (2)$$

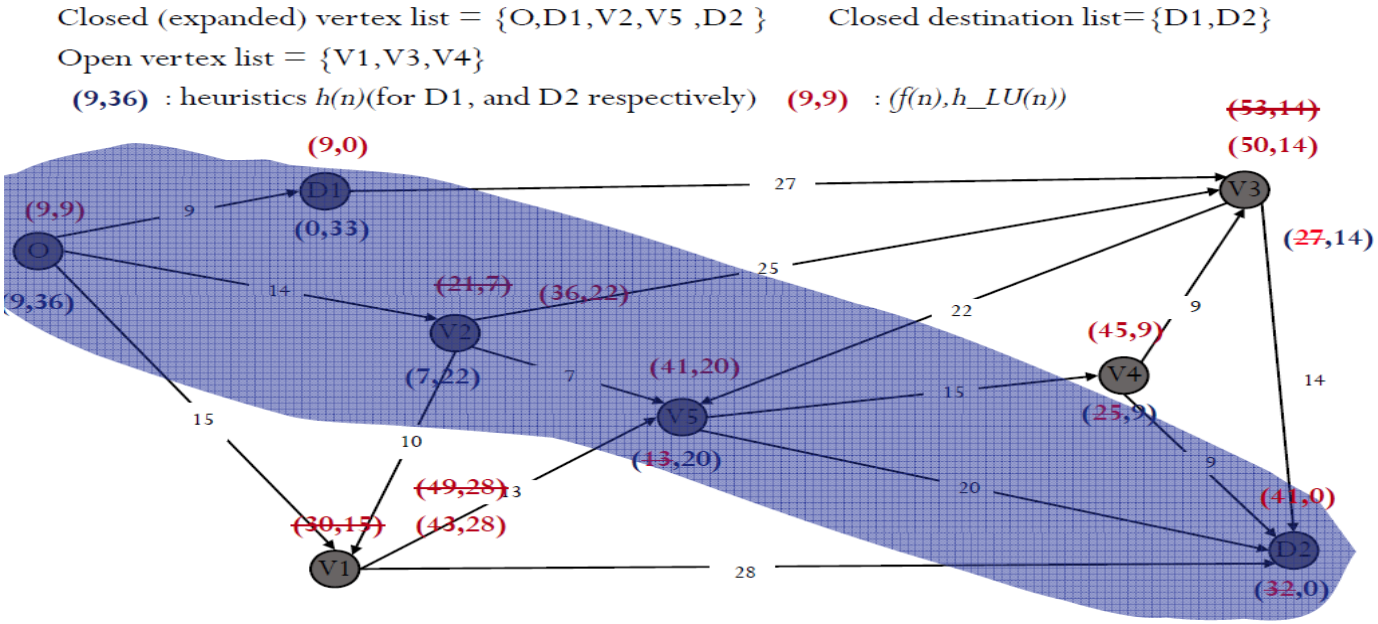


Figure 1. The LU search Process

where K is the number of unclosed, i.e., unreached, destination, and

$h_i(n)$ is the heuristic between n and the ith unclosed destination.

LU incrementally searches all paths leading from the starting vertex until it finds the path of minimum cost to any destination. It first takes the partial paths most likely to lead towards the unclosed destination appearing to have the least cost.

Similar to A*, LU also maintains a *closed list* where expanded states are stored, and a set of *partial paths*, unexpanded leaf states of expanded states. These partial paths are stored in an *open list*, also called a *priority queue*, a modified queue that outputs the one with the highest priority. Differently from existing best first searches, 1) LU has an array that marks a destination as closed after its path is found by LU, and a destination is *unclosed*, or *open*, when a route for the destination is not found yet; and 2) one additional variable, *Dest_H*, is adopted as one component of a generated state, *gs*, to indicate the unclosed destination appearing to be the one to which *gs* is the closest. In other words, *Dest_H* is used to indicate the unclosed destination towards which the heuristic is equal to *h_LU(gs)*.

Whenever an equal *f(n)* occurs, one is randomly selected to expand.

To retrieve least-cost routes for N destinations, LU first generates the origin and puts it into the open list. Next, for all the states in the open list, LU expands the state with the lowest *f(n)* value, and its children states are generated and their f values are evaluated based on all the unclosed destinations. The process continues until all destination states are reached or no solution is found, i.e., at least no route for one destination is found. Once a destination, *DS*, is reached, the algorithm will output the obtained path, mark the destination as closed, reevaluate the generated but unexpanded states whose *Dest_H* is equal to *DS*, and update their *f(n)*s in the open list by removing *DS* from consideration to calculate *h_LU(n)*, which results in equal or larger *f(n)*s.

LU is optimal, i.e., the retrieved routes are all optimal, when $h_i(n)$ is never larger than its corresponding actual cost, i.e., the actual least-cost between n and the ith unclosed destination.

B. An Example

Figure 1 presents the state snapshot when LU finishes its search for a problem asking for shortest routes from O to D_1 and D_2 respectively. For each vertex, ($cost_1$,$cost_2$) in blue represents the cost estimations, or heuristics, where $cost_1$ represents the estimated cost, or heuristic, to D_1 and $cost_2$ represents the estimated cost to D_2. For a vertex n, ($cost_3$,$cost_4$) in dark red represents the cost estimations where $cost_3$ represents f(n), and $cost_4$ represents *h_LU(n)*.The cross line in ~~($cost_3$,$cost_4$)~~ indicates that a state is generated first and then either discarded directly or put into the open list and later removed from the open list.

The search starts with the generation of state O in the open list. Its cost to D_1 and $_{D2}$ are evaluated and the corresponding *f(O)* and *h_LU(O)* are calculated. Next, since O is the only state in the open list, O is expanded, and its three children states, V_1, V_2, and D_1, are generated and their costs to the destinations are evaluated. Now D_1 has a minimum *f(n)* value, so it is expanded. A state for V_3 is generated and put into the open list. Since D_1 is a destination, D_1 is put into the closed destination list, and the state for V_1 and V_2 are re-evaluated only based on the cost estimation to D_2. Accordingly, *(30,13)* for V_1 is changed to *(43,28)*, and *(21,7)* for V_2 is changed to *(36,22)*. Then V_2 is selected because it has a smallest *f(n)* among V_1, V_2, and V_3. It generates its children states for V_1, V_3, and V_5 and put them

into the open list. Since the original states for V_1 and V_3 have smaller *f(n)*s, their newly generated states are discarded. Next, since V_5 has the lowest *f(n)* in the open list, it is expanded, and its children states for V_4 and D_2 are generated and put into the open list. Next, since D_2 has the lowest *f(n)*, it is expanded. Since it is a destination, it is put into the closed list. Now the closed list contains all the destinations, so the search stops. Therefore, the optimal routes obtained are $O->_{D1}$, and $O->V_2->V_5->D_2$.

C. Discussion

LU can be used for both route planning and dynamic routing. When a traffic event occurs, the costs of the corresponding street links can be adjusted accordingly. For example, the costs for closed streets can be considered as infinite and for congested links can be high so that an alternative route not traversing the closed streets can be chosen as the best route. Under these situations, only the graph representing the street network needs update, and no need to change the algorithm LU.

Compared to Dijkstra's algorithm, LU gains performance through reduced state generations. However, to retrieve *N*-destination least-cost paths, LU may not outperform *N* sequentially-running A*s when their corresponding expanded states do not significantly overlap. The major reason is that compared to an A* search process, LU must maintain a longer closed list and a longer open list. Consequently, it requires more time to update and re-order the open list and search the closed list. This additional cost will be dominant if the expanded states between different A* processes do not significantly overlap. Therefore, when the destinations are far away from the origin and not clustered, for example, uniformly distributed in a transportation network, LU may not outperform A*. When carpooling, generally workers are both living closely and working closely. In other words, both the origins and the destinations are likely to cluster, which indicates LU may perform much better than multiple sequentially-running A*s.

Since a pickup location, represented by *pkloc*, for a person is always traversed before his/her drop-off location, represented by *drl*oc, during the least-cost route calculation, no need to calculate any route from *drloc* to *pkloc*. In LU and Dijkstra's algorithm, this can be achieved by neglecting the corresponding *pkloc* as a destination when a route starts with a *drloc*.

IV. EXPERIMENT AND RESULT ANALYSIS

To investigate the performance of LU for route pre-processing to retrieve a carpool-based least-cost route traversing a set of pickup and drop-off locations, a set of experiments is performed, and Dijkstra's algorithm and multiple sequentially-running A*s are used as the baselines. Their performance is studied using network distance in a large dense urban transportation road network. In the experiment, each problem sample is to ask for a set of shortest routes, each of which is between two locations within the pickup homes and drop-off workplaces.

The Euclidean distance is used as the basis to calculate the heuristic *h_LU(n)* for each generated vertex *n* in LU. Since a Euclidean distance between two vertices in transportation network is never larger than the actual network distance, LU is optimal.

The experiment uses one large dense urban transportation network, the road network of Fairfax City and Fairfax County, US that contains 35,435 vertices and 82,926 directed edges, as shown in Figure 2. Both origins and destinations are clustered. In practice people use a van or a car for carpooling, so the number of pickup and drop-off locations, *N*, may not be larger than 12.

Figure 2. The road network of Fairfax county and Fairfax city, VA, US

Three data sets are generated. As shown in Figure 3, in each data set, an origin, represented by a green dot in Figure 3, is generated first, and then a destination around a specified Euclidean distance, *ED* (in mile), is generated. Thereafter, the other origins, represented by green dots in Figure 3, within a selected radius, *R*, of the origin and the other destinations, represented by red dots in Figure 3, within the same radius of the destination are generated. The origin number is either equal to or 1 larger than the destination number.

Data set I is used to investigate the impact of the number of pickup and drop-off locations on LU, Dijkstra's algorithm, and A*. It varies *N* from 3 to 12. *ED* is 10 miles and *R* is 0.5 mile. For each *N*, the number of problem samples, *PS*, is 30.

 ISBN: 978-1-61208-118-2

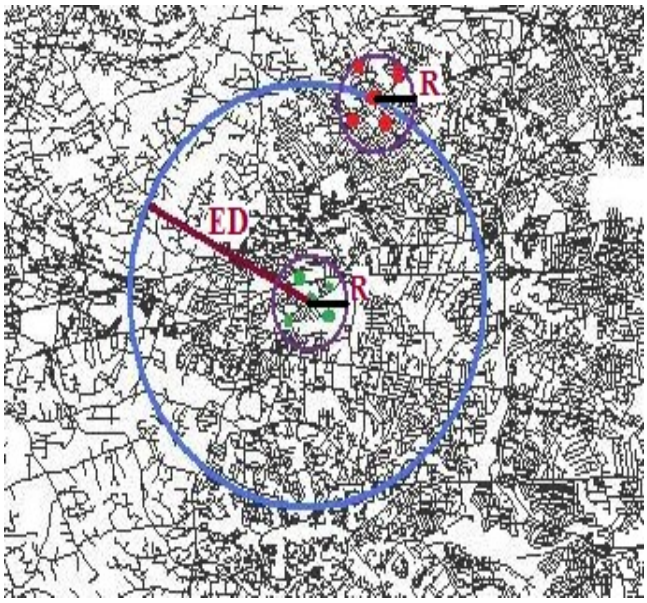


Figure 3. Data set generation

Data set II is generated to investigate the impact of the distance between homes and workplaces on LU and A*. It varies *ED* from 8 miles to 18 miles, with a fixed 2-mile interval. *N* is 3, and *R* is 0.5 mile. For each *ED*, *PS* is 30.

Data set III is adopted to investigate the impact of the radius size for carpool on LU and A*. It varies *R* from 0.5 mile to 1.0 mile. *ED* is 10 miles and *N* is 3. For each *R*, *PS* is 30.

A*, Dijkstra's algorithm, and LU are implemented with C#. The experiments are performed on a Toshiba Satellite A215 Laptop with 2.0GB memory (RAM), AMD Turion™ 64*2 Mobile Technology TL-56 1.80HZ processors, and Windows Vista™ Home Premium operating system.

A. Performance Measures

The following measures are used to analyze the performance of LU, Dijkstra's algorithm, and A*.

Average Shortest Distance (ASD): the average sum of shortest route distances obtained over all runs, in mile;

Average Number of States Expanded (ANSE): the average number of expanded states obtained over all runs;

Maximum Additional Number of States expanded by Dijkstra's algorithm or A (MaxANS):* For each run, compared to LU, obtain the additional number of states expanded by Dijkstra's algorithm or A*, and the measure is the maximum among all runs;

Minimum Additional Number of States expanded by Dijkstra's algorithm or A (MinANS):* For each run, compared to LU, obtain the additional number of states expanded by Dijkstra's algorithm or A*, and the measure is the minimum among all runs;

Average Process Time (APT): the time required to return the solution for a query, in second;

Maximum Additional Cost by Dijkstra's algorithm or A (MaxAC):* For each run, compared to LU, obtain the additional time cost required by Dijkstra's algorithm or A*, and the measure is the maximum among all runs;

Minimum Additional Cost by Dijkstra's algorithm or A (MinAC):* For each run, compared to LU, obtain the additional time cost required by Dijkstra's algorithm or A*, and the measure is the maximum among all runs;

Average Relative Number of States expanded (ARNS): the ratio of the number of states expanded by Dijkstra's algorithm or A* over by LU; and

Average Relative Process Time (ARPT): the ratio of the time processed by Dijkstra's algorithm or A* over by LU.

B. Results

The results are provided in Table I through Table VI. Dijk represents Dijkstra's algorithm. The minimum and maximum are highlighted in bold for *ASD, MaxAC, MinAC, ARPT, MaxANS, MinANS,* and *ARNS*. "-" represents a value is not available due to the high computation cost.

It is observed that all ASDs obtained from LU are the same as from Dijkstra's algorithm and from A*, which is direct evidence showing that LU retrieves optimal solutions with Euclidean distance as the basis to calculate its heuristics.

TABLE I. THE PERFORMANCE ON AVERAGE SHORTEST DISTANCE AND PROCESS TIME FOR DATA SET I

N	ASD	APT			ARPT	
		LU	*A**	*Dijk*	*A**	*Dijk*
3	49.5	12.1	13.1	733.4	1.1	60.6
4	100.4	26.5	37.9	1060.1	1.4	40.0
5	153.8	38.3	66.7	1567.7	1.7	41.2
6	223.6	16.9	39.9	1750.7	2.4	103.5
7	313.3	48.9	118.0	1746.4	2.4	35.7
8	394.9	36.5	75.1	2093.2	2.1	57.3
9	438.8	39.1	97.8	-	2.5	-
10	640.9	45.6	171.4	-	3.8	-
11	775.9	59.7	205.4	-	3.4	-
12	931.8	73.0	264.7	-	3.6	-

TABLE II. THE PERFORMANCE ON EXPANDED STATES FOR DATA SET I

N	ASD	ANSE			ARNS	
		LU	*A**	*Dijk*	*A**	*Dijk*
3	49.5	9972	76348	59372	7.6	5.9
4	100.4	14802	226009	82399	15.2	5.5
5	153.8	20328	593254	109821	29.1	5.4
6	223.6	19649	952359	124700	48.4	6.3
7	313.3	31809	2544339	133529	79.9	4.2
8	394.9	28206	2633455	150217	93.3	5.3
9	438.8	25504	5893107	-	231.0	-
10	640.9	43738	9211019	-	210.5	-
11	775.9	51828	14425335	-	278.3	-
12	931.8	57312	19725022	-	344.1	-

 ISBN: 978-1-61208-118-2

TABLE III. THE PERFORMANCE ON AVERAGE SHORTEST DISTANCE AND PROCESS TIME FOR DATA SET II

ED	ASD	APT		MaxAC	MinAC	ARPT
		LU	*A**			
8	40.5	4.2	5.2	1.8	-0.1	1.2
10	49.5	12.1	13.0	4.9	-1.2	1.1
12	59.4	13.1	15.4	5.9	-0.3	1.2
14	64.5	19.9	23.2	9.8	-0.5	1.2
16	70.1	96.9	146.8	246.9	0.4	1.5
18	83.3	110.9	139.8	119.6	-15.1	1.3

TABLE IV. THE PERFORMANCE ON EXPANDED STATES FOR DATA SET II

ED	ANSE		MaxANS	MinANS	ARNS
	LU	*A**			
8	6587	51362	59519	31877	7.8
10	9972	76348	126260	26931	7.7
12	12438	97819	132983	52870	7.9
14	14863	114118	143007	50346	7.7
16	21728	206193	301551	43721	9.5
18	24109	196415	294018	87167	8.1

TABLE V. THE PERFORMANCE ON AVERAGE SHORTEST DISTANCE AND PROCESS TIME FOR DATA SET III

R	ASD	APT		MaxAC	MinAC	ARPT
		LU	*A**			
0.5	49.5	12.1	13.0	4.9	-1.2	1.1
0.6	50.5	14.2	17.7	10.2	0.7	1.3
0.7	50.0	11.7	13.7	2.8	1.2	1.2
0.8	50.1	11.7	14.8	4.8	2.2	1.3
0.9	49.6	12.1	13.1	4.9	-1.2	1.1
1.0	49.7	11.8	14.5	5.0	1.1	1.2

TABLE VI. THE PERFORMANCE ON EXPANDED STATES FOR DATA SET III

R	ANSE		MaxANS	MinANS	ARNS
	LU	*A**			
0.5	9972	76348	126260	26931	7.7
0.6	11736	96549	112879	48974	8.2
0.7	11261	91033	99623	54530	8.1
0.8	11174	91257	99623	54530	8.2
0.9	9972	76348	126260	26931	7.7
1.0	10950	89573	99623	54530	8.2

Based on Table I through Table VI, the following conclusions can be drawn.

Compared to A* and Dijkstra's algorithm, based on MinANS values in Table II, Table IV, and Table VI, it is clear that LU always expands the least number of states. This is because LU is more informed than Dijkstra's algorithm and do not have to re-expand states, which occur when multiple sequentially-running A*s are used to retrieve all shortest pair-wise distances.

In practice, for carpooling, both pickup locations and drop-off locations likely cluster to reduce trip cost, gasoline usage, and emission. Compared to of A*, According to the ARPT values in Table I, Table III, and Table V, on average LU is about 0.1 to 2.8 times faster than A*. Especially, based on ARPT in Table I, when *N* increases, LU is increasingly faster than A*. According to ARPT in Table 1, it is clear that when the number of pickup and drop-off locations increase, Lu increasingly outperforms A* because more states are re-expanded by sequentially-running A*s.

It is clear that when *N* is small, LU significantly outperforms Dijkstra's algorithm in terms of computation efficiency. Based on Dijkstra's ARPT values in Table I, LU can outperform Dijkstra's algorithm by 2 magnitudes.

Based on ARPT values in Table III, when *ED* increases, generally LU is increasingly faster than A*.

Based on ARPT values in Table V, compared to of A*, the performance of LU does not have a clear relation to *R*.

Based on MinAC values in Table III and Table V, in some rare cases, LU may still be less efficient than A*.

V. CONCLUSION

Within a location based social network, carpooling is becoming more and more preferable among workers both living and working near each other due to the continuous increase in gasoline price and air pollution, and is more desirable when working places are far away from homes. Consequently, it is highly desirable to obtain the optimal traversal order to pick up carpool participants and drop off them to retrieve the least-cost carpooling route. However, in a large network, it is desirable to first compute least-cost pair-wise distances among the pickup and drop-off locations to reduce the computation complexity to retrieve the optimal route for carpooling.

In this paper, LU, a fundamental best first search algorithm and framework, is adopted to pre-compute all least-cost pairwise routes. In a carpooling scenario, both pickup locations and drop-off locations are likely to cluster to get most out of carpooling in terms of reductions in trip cost, emission, and gasoline usage. Accordingly, compared to the two existing prevalent algorithms, A* and Dijkstra's algorithm, LU is more appropriate to compute all least-cost pairwise network distances. A set of experiments is performed, and the results demonstrate that LU expands the least number of states when compared to A* and Dijkstra's algorithm, and 2) on average LU is much more efficient than A* and significantly faster than Dijkstra's algorithm when the number of pickup and drop-off locations are not larger than 12. On average, LU is 0.1~2.8 times faster than A* and outperforms Dijkstra's algorithm by 2 magnitudes.

VI. FUTURE RESEARCH

Even though LU significantly reduces overlapped states expanded by multiple sequentially-running A*s, LU may still be less efficient. One major reason is that unnecessary states having been used to search for the routes to the closed destinations but not helpful for searching the routes to the

 ISBN: 978-1-61208-118-2

remaining unclosed destinations are not removed timely in the current implementation of LU. Future research can be performed to reduce unnecessary states stored in the open list and the closed list whenever a destination is closed to expedite the search and update operations performed on both lists in LU, and thus to further improve the efficiency of LU.

REFERENCES

[1] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook. The Traveling Salesman Problem: A Computational Study. Springer, 2007.

[2] S. Arora. Approximation Schemes for NP-hard Geometric Optimization Problems: A survey. Mathematical Programming, Springer, 97 (2003) pp. 43-69.

[3] E. W. Dijkstra: A Note on Two Problems in Connexion with graphs. In Numerische Mathematik, 1 (1959), S. pp. 269–271.

[4] Richard Bellman. On a Routing Problem, in Quarterly of Applied Mathematics, 16(1), pp. 87-90, 1958.

[5] Q. Lu. LU: A Best First Search to Process Single-Origin Multiple-Destination Route Query in a Graph. Proceedings of the 2010 Second International Conference on Advanced Geographic Information Systems, Applications, and Services, (2010) pp. 137-142.

[6] P. E. Hart, N. J. Nilsson, and B. Raphael. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. IEEE Transactions on Systems Science and Cybernetics SSC4 (2) (1968) pp. 100–107.

[7] Grout, V., (2003), Towards an Optimal Routing Strategy, Proceedings of IADIS WWW/Internet 2003, Algarve, Portugal, 5 th. - 8th. November, pp. 903-906

[8] Pierre A. Humblet. An adaptive distributed dijkstra shortest path algorithm. Technical Report CICS-P-60, Center for Intelligent Control Systems, MIT, May 1988

[9] Baruch Awerbuch. Shortest Paths and Loop-Free Routing in Dynamic Networks. SIGCOMM '90 Proceedings of the ACM symposium on Communications architectures & protocols: pp. 177-187

[1] Dafermos, Stella. C. and F.T. Sparrow. The Traffic Assignment Problem for a General Network." J. of Res. of the National Bureau of Standards, 73B, pp. 91-118. 1969.

 ISBN: 978-1-61208-118-2

A Heuristic Approach Based on the Ant Colony Optimization for the Routes Elaboration on the Fuel Collection for the Brazilian Petrolium Agency

Alex Barradas, Adriano dos Santos, Sofiani Labidi, Nilson Costa
Federal University of Maranhão
São Luís, Brazil
{barradas.alex,adriano.asr,soflabidi,nilson2001}@gmail.com

***Abstract* - The trade of petrol derivatives such as natural gas and biofuel is an activity that demands supervision and monitoring. The ANP (Petrol, Natural Gas and Biofuel Brazilian Agency) is responsible for guaranteeing the quality of the products made by the petrol industry. Nevertheless, in order to be efficient, the ANP has established a partnership that authorizes the LAPQAP/ UFMA (Analysis and Research in Petrol Analytical Chemistry Laboratory/Federal University of Maranhão) to be its representative in the State of Maranhão. The proposal of this research is to present a prototype in order to automatize part of the fuel sample's collection process in order to make the tasks simpler, which nowadays are executed by the lab's analysts, who make that work by their experiences, increasing the number of supervised fuel stations aiming to diminish the irregularities on fuels traded in the State of Maranhão. Developed the prototype called S-Rota that is responsible for creating a circuit between the starting point UFMA and the fuel stations that approaches the concepts regarding the Ant Colony Optimization Algorithm (ACO).**

Keywords - Collection; Monitoring; Quality; Fuel; ANP; ACO.

I. INTRODUCTION

The ANP is responsible for the execution of the national policy for the Petrol, Natural Gas and Biofuel Power Sector according to the Petrol Law (Law number 9.478 / 1997). To make these actions concrete, the ANP has developed a Fuel Quality's Monitoring Program (PMQC) and established a representative in each state.

The PMQC has as objective to systematically evaluate the quality of the fuel traded in Brazil (Gas, Diesel, Ethanol and B2 mix) mapping situations that are good enough according to its parameters so that supervising actions can be directed. The PMQC has 23 partner laboratories.

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 Maranhão State, as well as sending Brazilian Automobiles Liquid Fuels' Quality monthly reports to ANP.

Thus, the program developed by the ANP, the PMQC needs to be analyzed, identified and characterized in its main stages so that it can be contextualized to Maranhão's scenery. Afterwards, it is observed the Fuel Sample's Collection (CAC) and its relationship with the other stages executed by the LAPQAP.

At the Fuel Sample's Collection stage, the LAPQAP focus on evaluating the costs and time spent for the inspected fuel stations definition, routes creation and samples' collection. These days, the elaboration of the definition of the fuel stations inspected and the route creation is handmade. This action puts all the responsibility on the collectors (driver and analyst). This stage is fundamental because it influences the conclusion of all the other stages of PMQC and is subject to human mistakes.

The developed prototype elaborates the better route between the LAPQAP and the reseller fuel stations, having this approach based on the Ant Colony Optimization [1].

II. PMQC

The PMQC concept is to ascertain the quality's standard of the traded fuel on its extraction, refining, distribution and reselling stages to the national market [2]. PMQC's main objectives are surveying the general indicators of the fuel's quality traded in the country and the identification of non-conformities focus, aiming to orientate and improve the Agency's inspection area performance [2].

A. *PMQC stages*

The three stages that compound PMQC are:

- Fuel Samples' Collection: The first stage is based on the accomplishment of raffles that indicates the stations that will be inspected and the fuel samples' collection by the LAPQAP.
- Samples' Lab Analysis: The second stage is responsible for collecting the samples that are conforming the ANP regulation and analyze their quality.
- Data Treatment and Information sending to ANP: This stage finds the result generated on the second stage to organize and insert them on the system of Fuel's Quality Monitoring (MQC) to send them to ANP after all.

B. *Fuel Samples' Collection Stage (CAC)*

This work objective is focused on this stage, aiming to automatize part of the collection process. In State of Maranhão, the existing fuel stations (can be called universe) is saved on the data base lab which represents the ANP in the State.

 ISBN: 978-1-61208-118-2

Besides that, aiming to make the logistics easier, the Maranhão State is divided in four regions by the LAPQAP called R1, R2, R3 and R4, as shown in Figure 1. This way, the laboratory has one week to collect the samples in each region.

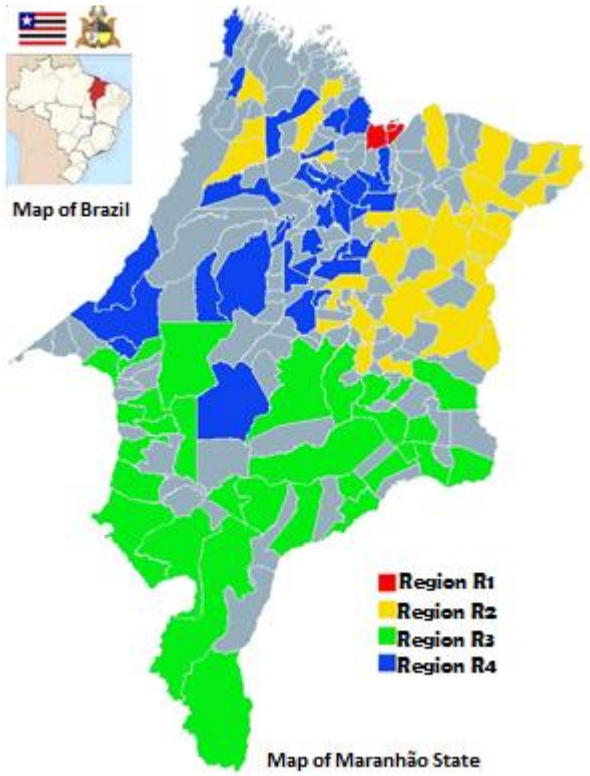


Figure 1. Four Regions [1].

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 .

The sampling in this work focus in the region 1, considering no have differences in generation routes among the regions.

Then, when the fuel stations to be inspected are defined, the responsible collaborator will have the route between the starting point – UFMA and the fuel stations.

The routes nowadays are done without the help of any tools and/or logistical techniques, only the collaborator's empirical knowledge, even though the LAPQAP establish some criteria to prioritize some fuel stations above others.

The main criteria are divided in:

- Fuel's Quality.
- Documentation.
- Distance.

For the fuel's quality criteria, the last sample of the three types of fuel: alcohol, gas and diesel. According to the result of each type of fuel, a level of priority is established.

For the documentation criteria, the LAPQAP requires the inspected level to present the documentation required by the ANP. Then, the laboratory elaborates a list that discriminates and informs the situation of each document.

For the distance criterion, the region to be inspected is verified and the shortest distance between the starting point and the fuel stations are is analyzed. In other words, the closest fuel stations to the starting point are prioritized. After this, the analyst evaluates the three criteria and defines the order of fuel stations to be inspected. It is important to mention that each criterion is applicable only to 10% (ten percent) of the make fuel stations.

III. ANT COLONY OPTIMIZATION

The Ant Colony Optimization (ACO) [3] is based on real and solved problems by an ants' colony. This algorithm is a heuristic based on probability, created for solution of computational problems that involve search of ways in graphs. The ants' colonies are distributed systems that despite the individual simplicity represented in an isolated manner, the system's structure present a high leveled social organization [4].

The main functions executed by an ants' colony between the relationships among the ants include: building or increasing the nest, finding food, feeding the brood [5].

On the interactions between ants searching for food and building trails, a substance called pheromone is used. This substance influences the ants in the choice of innumerous routes, in other words, how much more the tax concentration of pheromone in a certain way, the biggest will be the chances for an ant to choose the same way [6].

The main characteristics are:

- The ants' population (agents) is an independent being moving around simultaneously without a central control.
- The search for routes happens in a deterministic way.
- The algorithm is cooperative, namely, each ant chooses a way based on the pheromone's concentration deposited by other ants.

The algorithm main idea is the indirect communication based on the pheromones routes between an agents' colony called ants [7]. The mains mathematical formulations that compound the algorithm are described below.

The equation (1) illustrates the mathematical formulation used on the work for the ants' colony algorithm [8] for indicate a probability of ant κ in this point i and choice point j:

 ISBN: 978-1-61208-118-2

$$p_{ij}^{k}(t) = \begin{cases} \frac{[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\Sigma[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}} \\ \kappa \in allowed_{\kappa} \\ 0, others\ cases \end{cases} \quad (1)$$

On the equation (1), the variable α is a weighting of the pheromone $(0 \leq \alpha \leq 1)$ and β is the a weighting of the heuristic information $(0 \leq \beta \leq 1)$, where $\eta_{ij} = {}^{1}/_{d_{ij}}$ is the visibility between the variables i and j.

The equation (2) shows the formulation that defines the pheromone deposited on the route [8]:

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L_k}, if\ the\ \kappa - th\ ant\ use\ the\ track\ (i,j) \\ 0, others\ cases \end{cases} \quad (2)$$

The variables that compound equation (2) are defined as follows:

- Q is a project constant.
- L_k is the length of the circuit of the κ-th ant.

When an ant finishes the circuit in a $t_0, t_0 + n$ time and consists in a cycle of n interactions, the equation result is used to update the amount of substance deposited on the route, based on equation (3) [8].

$$\tau_{ij}^{k}(t+n) = \rho \cdot \tau_{ij}^{k}(t) + \Delta\tau_{ij}^{k} \quad (3)$$

On equation (3), the variable ρ represents the route's persistence during the cycle $(0 \leq \rho \leq 1)$, in which the value $(1 - \rho)$ the trail evaporation between the field and the t and $t + n$.

This way, the algorithm ACO searches for a feasible solution for the famous Travelling Salesman Problem (TSP), it is a problem that belongs to the category NP-Hard from complexity exponential [4]. Traveler simulating an ants' walk by a graph using the mathematical modeling for the concentration of pheromone in the graph [9]. However, the ACO algorithm does not guarantee the best way but a good solution with polynomial execution cost.

IV. RELATED WORK

The TSP is of importance since similar combinatorial optimization problems arise in industrial applications and can be formulated as TSP-like instances [10].

In the business scenario, Google invests in solutions to the TSP through the routes services available in the Google Maps API V3. By default, the routes service calculates a path which includes the reference points provided in the indicated order. Optionally, the routes service may optimize the path provided by rearranging the points of reference in a more effective order [11].

However, the routes service offered free by Google Maps API V3, have some limitations in regard the amount of allowed points restricting to a total of eight points. Moreover, the API does not offers directly the insert of new criterions for the creation of the path.

The application S-route uses the Google Maps API only in georeferencing services and applies the ACO approach for route optimization that allows the use of up to 60 tested points. Another S-Route differential, is presente in the possibility of developing routes based on criterions such as fuel quality and documentation**.**

V. PROPOSED SYSTEM

The prototype s-Rota aims to automatize part of the fuel samples' collection process of LAPQAP. So, the procedures executed by the LAPQAP are analyzed aiming to absorb and understand its main features.

A. S-Rota modeling

In order to make the visualization of the activities' diagram from the Figure 2, the activities' diagram shown in Figure 2 has been used. This activities' diagram is a generalization.

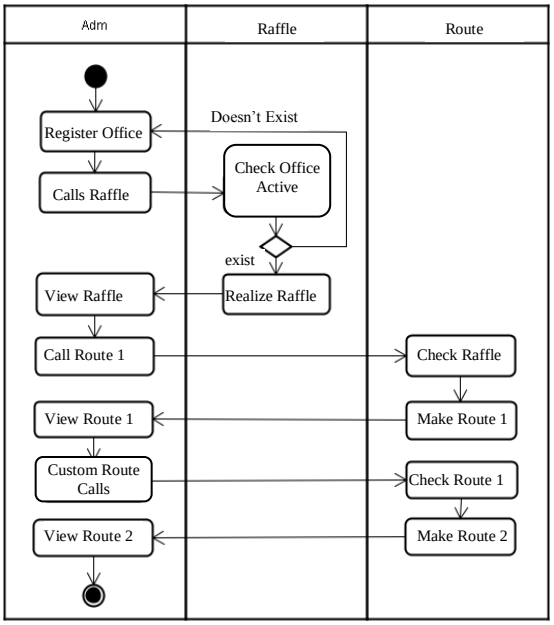


Figure 2. Activities' Diagram.

Figure 2 is a generalization, called, a general summary of the routes' creation process by the prototype. The stage starts with the administrator enrolling the existing fuel stations and requiring the lottery or assortment in execution.

Before that phase, the existence of stations in the data base is checked; if there are stations, the raffle is done and sent to the administrator. Then the routes definition process is started by the user.

After the prototype search for the last raffle done and elaborates the route based on the distance called rota1, that is, at this moment the shorter a total route is, the better. After the router 1 is elaborated, the administrator asks for the elaboration of personalized route called router 2. The activities' diagram finishes with the visualization of router 2 by the administrator.

 ISBN: 978-1-61208-118-2

B. S-Rota Implementation

The S-Rota prototype has been implemented with the composition of the following elements:

- Web Server / S-Rota web application.
- Maps API .
- Data base manager system (SGBD).

The web server is responsible for storing and make it available on the net the S-Rota application, specifically on the Internet. The software is accessed as any conventional website. The languages used for the prototype of the application making were PHP [12] and Java Script [13].

For the maps API, it has been searched an API of maps that offers services dynamically geo-referred and without costs concerning software's license. An option to the approached criterion is Google Maps [11].

As the maps API uses Java Script, the applied searching algorithm has been developed in this language, to make the communication between the maps API and the prototype that uses data structure in XML format as well easier [14]. Besides the programming languages, the prototype stores the information recorded and generated by the application in a relational SGBD.

To manage the objective of each element listed by the LAPQAP, the project has defined the following development stages:

- Data maintenance.
- Fuel Stations Selection.
- Routes elaboration.

C. Data Keeping

This stage represents the data management inserted at the S-Rota. Such activities are related to the administrator profile. The main items in this process are:

- Fuel station.
- Collector.
- Note.

In the S-Rota prototype, the information regarding the item fuel station stores in the table called Fuel Station. These information work as feeding for the next stages of Fuel Stations selection and Routes elaboration.

Besides that, the table Fuel Station is related to other tables such as the fuel's quality (qualitycomb) and the table documentation (documentation). Through these relationships it is possible to associate the fuel station to the quality of each fuel traded and the documentation historical; such information are essential to the elaboration of the personalized route.

The next item, called collector, represents the responsible for collecting the fuel samples. The data of each responsible collector are stored in a table called collector; such information contains personal data and the historical of each collector regarding the each collection done.

In the last item (notes), the administrator inserts in the data bank any unexpected event in the collection process, in other word, in the table called reminder it is registered any unwanted event.

D. Fuel Stations' Selection

At this stage, the prototype should select in a random and transparent way the reseller fuels stations that will be inspected by the LAPQAP.

In this process, a class called SORTEIOCLASS has been projected with several functions, among them it is cited the function sorteio() (raffle). This function basically checks the data bank of the active fuel stations and then, the function array_rand [8], the fuel stations are organized randomly conforming the limited amount of fuel stations defined by the function countSorteio().

By the user watching, the administrator access the raffle screen and asks for the Fuel stations selection. Then, a file in XML format containing the information about the raffle fuel stations needed for the maps' generation is generated.

E. Routes Elaboration

For the routes elaboration, it is necessary in advance to have a map with the fuel station selected. The Google Maps' carry the API file having all the necessary information for the localization and identification of each fuel station and the starting point (UFMA). With these steps, the S-Rota is able to represent each item in the map generated by the API, finding them in the Maranhão State.

Figure 3 shows the elements geo-referred, such as: starting point and fuel stations present in the XML file.

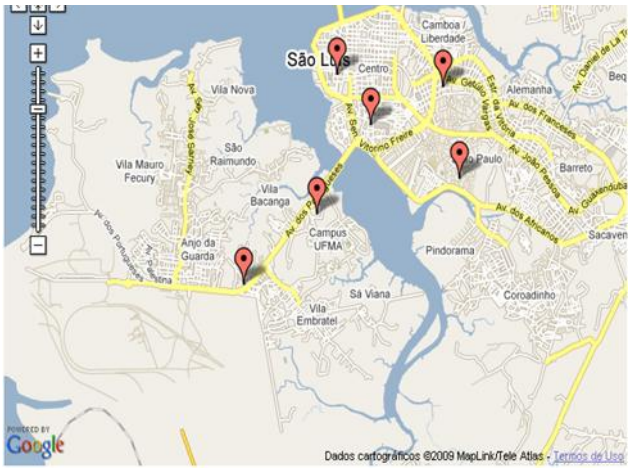


Figure 3. Route's elaboration.

The API acts basically in the representation of geographical information and in the geo-referred data input for the search algorithm.

The responsibility of evaluating the better way is given to the search algorithm. In this case, the algorithm optimization applied to solve the Travelling Salesman Problem had as a base the ant's colony.

In this context, the mathematical principles and the concepts regarding the ant's colony have been implemented in Javascript aiming to interact with the Google Maps' API. The number of ants (n Ants) and of interactions used by the algorithm has been 10 each [4].

 ISBN: 978-1-61208-118-2

In the following pseudo-code, the main function of the implementation of the ants' colony, the function ACO ( ) is represented.

```
function ACO(){
    Start the matrix of pheromone ;
    distAnts(nAnts); //'distributes the ants
    for i =0 to nIterations-1 then{
        for j=0 to nStation-1 then{
            for k=0 to nAnts-1 then{
                jSolution = generateRoutes(j);
                costSol=checkCost(jSolution);
                if(costSol < smallCost) then{
                    smallCost=costSol;
                }
            }
        }
        update_of_pheromone();
    }
```

In the pseudo-code, the function *generateRoutes(j)* represents the generation of the route executed by the ants which do by choosing the best points to follow the city *i* to *j* constructing the route according to equation (1). The function *update_of_pheromone()* represents the update of the matrix of pheromone that is carried through to each interactions *nInterations*, in which if increases or decreases the variable τ_{ij} (nivel of pheromone between *i* and *j* according to equation (3).

The following stage is the definition of the course of the personalized route. The process that establishes the new sequence of priorities between the fuel stations, analyzing the distance, fuel's quality and documentation criterion, conforming cited previously.

So, the new scoring of the fuel stations is obtained by an average of the three criteria, conforming the equation (4):

$$\frac{D\times P_D + C\times P_C + D_o\times P_{Do}}{P_D + P_C + P_{Do}}, \quad (4)$$

In which the variables are:

- D is the scoring of the fuel stations based on the distance.
- P_D is the weight defined by the administrator for the distance.
- C is the scoring of the fuel station based on the fuel quality.
- P_C is the weight defined by the administrator for the fuel's quality.
- D_o is the scoring of the fuel station based on the documentation.
- P_{Do} is the weight defined by the administrator for the documentation.

The variable scoring D, C *and* D_o may vary between the values from 0 to 100. For the scoring of C, in which the fuel station is irregular, in only one type of fuel, the scoring received is 33, for two types 66 and for three types 100. For the scoring of D_o , either is 0 for the irregular fuel station or 100 for the regular fuel station.

The weights are defined by the administrator through the prototype. Thus, the user may establish a priority order for each criterion according to the administrator needs.

F. Results

To get the prototype validation process started S-Rota, the LAPQAP provide real data referring to the collection done on October 22nd, 2007. The data extraction happened by the application G7ToWin version A.00.183.

The application carries the data obtained by the GPS Garmin that furnishes the geo-referred coordinates of each collected fuel station. The fields used for the results comparison were Fuel Station, Latitude, Longitude and Date/Time of the Collection as per figure 4.

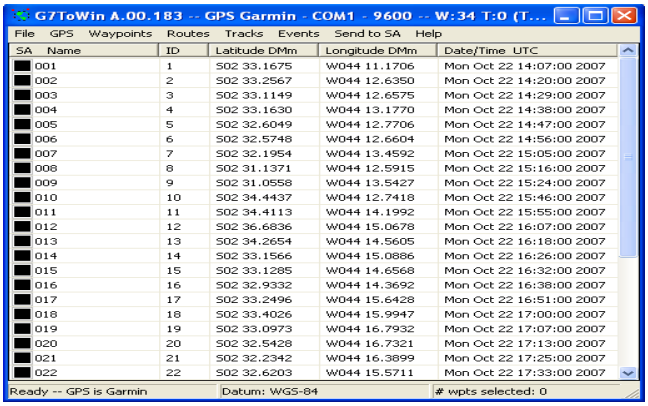


Figure 4. G7ToWin Data.

By the field Date/Time of the Collection it is possible to determine the sequence of visited fuel station in the time by the LAPQAP, which allows the route done by the driver to be traced. For the visualization of this course it was used the Google Maps, as per figure 5.

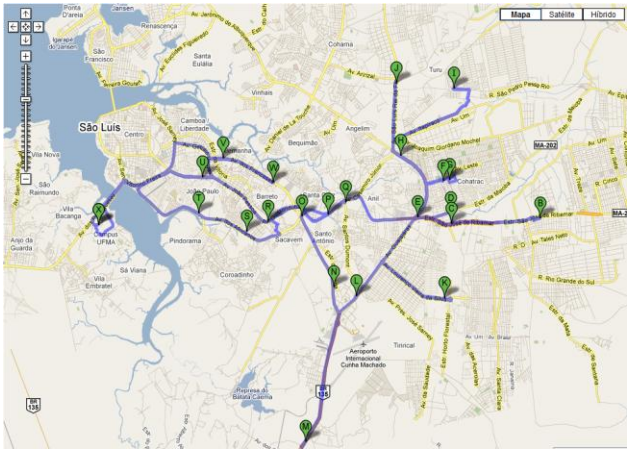


Figure 5. ANP Driver's route

The course done in figure 5 resulted in a total way of 85,49km. The elaboration of this route is done in an empirical way by the driver, not having a specific standard.

For the elaboration of the course by the prototype S-Rota, the following premises gave been considered:

- The variables $P_C = 0$ and $D_o = 0$.
- The same points of the last experiment as per figure 4.

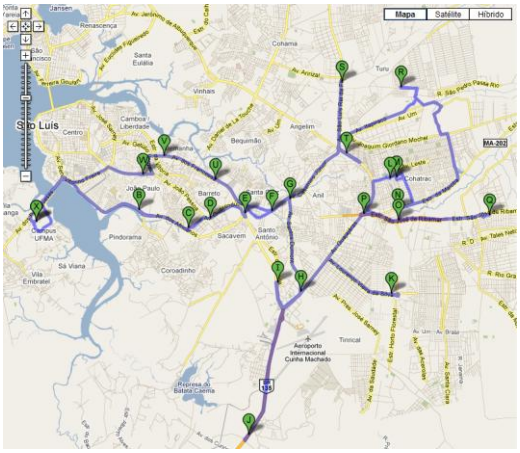


Figure 6. S-Rota's route.

The course elaborated by the prototype S-Rota resulted in a way of 77.48km. The difference between the figure 5 and the figure 6 is of 8.01 km.

The prototype accomplished the operation faster which has brought an economy of approximately 9.37% comparing to what accomplished the driver elaborating the route handmade.

VI. CONCLUSION

This research has presented a prototype that searches for the best way for the Fuel Samples' Collection stage. As a complement of this prototype called S-Rota, an optimization algorithm based on the optimization approach of ants' colony [7] has been developed, adapted to the language Javascript, applied to Google Maps' API and integrated to the S-Route base.

The prototype gathered with the optimization algorithm of ants' colony aims to automatize part of the process of Fuel Samples' Collection approaching concepts related to the UFMA's Analysis and Research in Petrol Analytical Chemistry Laboratory needs. It also creates the fuel stations in a random way, selecting a percentage by the user of active and existing fuel stations in the data base. This way, any inappropriate tendency in the fuel stations selection is avoided.

Despite the satisfactory and positive acquisition in this piece of work, some points still present limitations. As an example, the searching algorithm has not been tested in big fuel stations net yet and besides that does not evaluate the highways conditions and the traffic level. In these conditions, it is wished to improve the searching algorithm, in order to ascertain good results in big fuel stations nets.

This paper applied the Classic ACO for prove this code with minimal conditions its can resolved with success the problems best way for the S- Rota Software.

ACKNOWLEDGEMENT

Many thanks to everyone who helped directly and indirectly and allowed the creation of this software prototype.

REFERENCES

[1] A. Barradas, Abordagem Baseado na Heurística de Colônia de Formigas para Elaboração de Rotas na Fase de Coleta de Amostras de Combustíveis. São Luís, MA: Programa de Pós Graduação em Emgenharia de Eletricidade, Universidade Federal do Maranhão, 2009.

[2] ANP, "Programa de Monitoramento", Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, 2010. Accessed september 26, 2010. Available at http://www.anp.gov.br/?pg=33970.

[3] M. Dorigo, M. Birattari, and T. Stützle, "Ant Colony Optimization--Artificial Ants as a Computational Intelligence Technique," IEEE Computational Intelligence Magazine, 2006.

[4] M. Dorigo, "Ant Colony Optimization," MIT Press, 2004.

[5] K.F. Doerner, D. Merkle and T. Stützle, "Special Issue on Ant Colony Optimization", vol. 3, No. 1, Swarm Intelligence, March, 2009, pp. 1-2, doi:10.1007/s11721-008-0025-1.

[6] S. Nonsiri and S. Supratid, Modifying Ant Colony Optimization. IEEE Conference on Soft Computing in Industrial Applications (SMCia/08), Muroran, JAPAN, 2008.

[7] M. Dorigo, Optimization, Learning and Natural Algorithms., M.. PhD thesis, Italy: Politecnico di Milano, 1992.

[8] K. Kai, N. Haijiao, Z. Yuejie, and Z. Weicun, "Improved integrated optimization Model research of mode and route in multimodal transportation," IEEE International Conference on Information Management, Innovation Management and Industrial Engineering, 2009, doi: 10.1109/ICIII.2009.155.

[9] D. Angus and C. Woodward, "Multiple Objective Ant Colony Optimisation", vol. 3, No. 1, Swarm Intelligence, March, 2009, pp. 69-85, doi:10.1007/s11721-008-0022-4.

[10] B. Fox, W. Xiang, and H. P. Lee, "Industrial applications of the ant colony optimization algorithm," Int J Adv Manuf Technol, Published online: 6 January 2006, DOI 10.1007

[11] GOOGLE, API Google Maps, 2010. Accessed september 26, 2010. Available at http://code.google.com/intl/pt-BR/apis/maps/.

[12] PHP, Manual, 2010. Accessed september 23, 2010. Available at http://www.php.net/manual/pt_BR/preface.php.

[13] J. Resig, "Pro JavaScript Techniques", Apress, 2006.

[14] W3C, XML, 2010. Accessed september 26, 2010. Available at http://www.w3.org/standards/xml/.

 ISBN: 978-1-61208-118-2

Geographic Information System Models of 40-Year Spatial Development of Towns in the Czech Republic

Lena Halounová, Karel Vepřek, Martin Řehák
Dept. of Mapping and Cartography
Faculty of Civil Engineering, CTU in Prague
Prague, Czech Republic
e-mail: lena.halounova@fsv.cvut.cz
arch.veprek@tiscali.cz
martin.rehak.1@fsv.cvut.cz

Abstract—**There are many indicators of sustainable development of towns defined by urban specialists, sociologists, economists, etc. The paper presents the first part of a project whose goal is to find indicators of harmonic development of towns based on analysis of forty years development of fifty Czech towns. The indicators are studied in land use spatial changes, demography and road traffic intensity changes. First ten towns were processed for the period between 1970 and 2009 being mapped in general urban land use classes and related to the measured road density. City land use class areas were derived from combination of actual and historical city plans and remote sensing data using GIS tools. It was found that the traffic intensity within towns and to and from towns is more dependent on existence of close highways and by-pass roads unlike number of inhabitants, e.g. Political changes from the communist regime to the democratic one was also an important breakpoint in the city developments. Increase of the road traffic intensity and enlarging of residential areas are features proving the fact. The paper presents a methodology of spatial mapping of land use classes utilized for determination of town development. The town developments and their relation to road traffic is presented on maps and graphs.**

Keywords-GIS, remote sensing data, city plan, number of inhabitants, urban model, land use, road traffic intensity, coincidence table, multicorrelational analysis

I. INTRODUCTION

The development of cities during last decades has faced us with a new situation. Most inhabitants in many European countries are concentrated in large towns. One fifth of the Czech Republic population is living in three largest towns – Prague, Brno and Ostrava.

Present state of the balance among consumption level of society and quality of life is a matter of scientific papers, research [2], [3], [4], many projects [1], [7], and political and economical discussions in many countries. Life quality is directly related to a lot of environmental and socio-economic conditions. These conditions determine a harmonic development which should be based on equivalent and adequate demands of the human society. To define "adequate" means to take into account both consumption, and quality of life. Both are closely connected to the road traffic and its intensity.

The Department of Mapping and Cartography has been processing a project focused on a detailed evaluation of relations between the quality of life and present behavior of the human society. The project goal is to create a model allowing improving the present development status in urban areas being less demanding to ensure their sustainable development.

The project is a logical continuation of several projects performed by specialists from the Czech Technical University in Prague in the Czech Republic and the State Institute for Regional Planning who have collected and summarized large data volume of fifty towns (including three largest ones - Prague, Brno and Ostrava) on:

1. Functional typology comprising five classes housing and infrastructure areas, industrial and agricultural areas, areas of transport, areas of recreation including sport and green vegetation land, and areas of other functions,
2. Urban, agriculture, forest and water surface areas and other function areas from the Czech Office for Surveying, Mapping and Cadastre (COSMC) and Czech Statistical Office. The data are related to 1970, 1980, 1990, 2000, 2003, 2005 and 2008,
3. Basic components of the environment,
4. Basic components of the social and economical development.

The previous projects were focused on statistical data collected from the above mentioned sources and their processing. They did not comprise any spatial data and no spatial analyses were performed. Their city collection resulted in a large range of cities differing by size (from ten thousand to more than one million, by economic orientation (agricultural, industrial, university, touristic), by natural conditions (lowland surrounded by agricultural areas,

 ISBN: 978-1-61208-118-2

mountainous situated large forest areas), by geographic position in the republic – close/far to a frontier, etc. The city set is a good sample covering practically all Czech city types.

The processed project is focused on two new views – to select suitable indicators of the sustainable city development in the Czech Republic using also spatial characteristics together with already collected ones, and the role of the road transport intensity in the development.

land use city maps

Individual land use areas offer different conditions for living. The same land use classes in different areas and therefore all spatial units are characterized by a long list of attributes.

The spatial town development of first nine towns was derived from city plans designed by local administration, from aerial photographs collected between 1950 and 2008, and satellite image data (Thematic Mapper, MSS data) covering period between 1970 and 2000. The system of determination of individual time level of land use maps was based on map vector data representing municipal maps and the remote sensing data. Their content allowed to verify whether the mapped units are reality or the project author´s proposal. The supervised classification (maximum likelihood – used very often for land use/cover classification by many remote sensing specialists - [7], [8], e.g.) made the verification easier and quicker. The latest land use map was the first processed level and further step headed back to the previous levels. The land use maps were controlled by the statistical data of the Czech Statistical Office and Czech Office for Surveying, Mapping and Cadastre. Individual statuses for above mentioned years in urban land use functional classes – areas of housing, areas of facilities, industrialized areas, transport areas, recreational areas, green areas, and other areas.

II. ROAD TRAFFIC DATA

A large data base of the road network development has been created by the Road and Motorway Directorate. The data base comprises - among others - measurements of the road traffic intensity in many points of roads of various road classes since 60-ies (1968, 1973, 1980, 1990, 1995, 2000, 2005, and 2010) of the 20th century. The road traffic intensity is a number of vehicles. per 24 hours which passed through a determined point on a road in both directions. The measurements are collected in several thousands of selected locations in the Czech Republic and represent 24 hours´ period. It is an average of several 24 hours´ s data collection.

The measurements are available in map forms where each location is marked together with total amount of passed vehicles (including motorcycles), and tables where the amount is enumerated in a more detailed way distinguishing heavy-duty vehicles, cars and motorcycles.

III. SOCIO-ECONOMIC DATA

A deep analysis of another large data volume which has been collected since the second half of the 20th century will be performed in the proposed project. The data comprise 10-year research of socio-economic data of environmental changes performed at the University of Economics in Prague, e.g. Each city is described by several hundreds of statistical data. The data were collected by many students of the University within their thesis. The processing of the data is not presented in this paper and is a matter of the further research.

IV. CZECH TOWNS AND THEIR DEVELOPMENT

The Czech Republic does not have a continuous political and urban development. The development was formed mainly by political decisions having a decisive role of the urban land use changes. After the Second World War the urban development was relatively uneven and can be characterized by three types of cities. One type of cities had only a relatively slow and continuous spatial evolution within their administrative boundaries. The second type are cities with growing administrative areas; however, this growth was artificial as a result of political decisions to join surrounding villages to a close city. The third group of towns is similar to the second one; the only deference is in the further separation of one or more early joined villages. This development classification was firstly described and denoted by Vepřek [8]. He uses three new terms: core area for town size representing in most cases a status in 1970. It was a year when the process of joining villages to neighbor cities became an important phenomenon in administrative structure of the country. The joined areas are named associated areas by Vepřek in [8]. Urban parts in associated regions are called agglomerated ones. The parts which became independent villages later on, are called peripheral areas Vepřek in [8].

This spatial development is archived in Cadastre books in the form of table records showing concerned cadastre districts. The transfer of this information can show the spatial city development using the cadastre districts´ boundaries of an appropriate period. This transfer was performed also into city plans, whose processing intervals vary in individual towns. City size evolutions cannot be derived from remote sensing data. If a town belongs to the second or third group, there are large spatial changes. The largest parts of these changes are in prevailing part represented by agricultural areas. The main difference between a core area and associated area are separated urban parts occurring in them.

V. LAND USE CLASSES DETERMINATION

Land use class definitions were created from land use classes applied by urban engineers in city plans which were processed in several-year periods. Their classes in different cities are not standardized even though their dissimilarities in

 ISBN: 978-1-61208-118-2

various cities are not important. However, their classification is too detailed for a general land use evaluation and was simplified to seven or nine land use classes in urbanized areas which were finally reclassified to five functional urban classes. These were the city plans which were the first information layer for urban land use maps processing.

The first step of preparing urban land use classes is a reclassification of detailed legend classes. The final functional classes were residential, production, recreational, traffic and other areas. Each class is therefore formed by a higher number of city plan classes. The residential area is formed by mixed residential region, general residential and rural ones, and public areas. The reclassification means also including local roads belonging to roads of low level in the state roads hierarchy into residential or other surrounding areas. The reclassification is performed individually for each town according to its city plan classes.

The advantage of this approach was the fact that the basic classification was performed by urban specialists, however, city plans comprise not only a real status, but also an urban plan. Therefore the next step was to verify the present city plans and the real state of towns as the plans comprise plans which may and really significantly differ from the real state especially in newly urbanized areas. This part of the processing was done by visual interpretation of the remote sensing data (aerial photographs) combined with a change registration/vectorization of the vector city plan and the result was a map of functional classes of the present state.

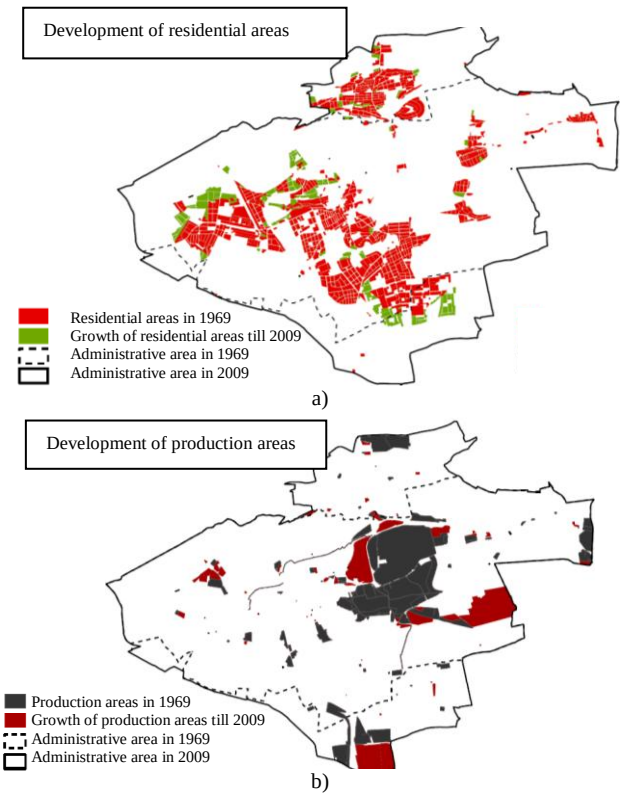
The forty years development was divided into determination of previous phase maps of functional classes in 2000, 1990, 1980 and 1970. The process started in the latest period and ended in 1970. These maps were processed using remote sensing data. Two types of remote sensing data were used for the change detection allowing mapping of the given year. The latest map (2008) was a result of the present city plan processing by implementing corrections found in discrepancies between the plan and aerial orthophotographs. Satellite data - Thematic Mapper (TM) data - were utilized for determination of land cover changes. They were derived from a subtraction of original satellite image bands and normalized vegetation index in two different years (2008 band minus 2000 band). Found changes were pixels with high positive or negative values. This approach yielded areas with different land cover, however, there was an additional task to determine and "translate" a land cover change into a land use change. Each functional class comprises a wide range of the land cover classes in the aerial photographs spatial resolution; however, these detailed classes are not in prevailing part detectable on the Thematic Mapper data. The Thematic Mapper resolution does not allow determining urban functional classes – agriculture area spectral behavior can be similar to vegetated areas for some plants, etc. The areas with important changes (extreme positive and negative pixel values) were verified using the aerial photograph taking into account also their shape and texture. The oldest map showing the 1970 year was also visually controlled using aerial orthophotomosaic created from aerial orthophotographs collected in 50-ies in the last century.

All functional classes were controlled by the statistical table data available at the Czech Office for Surveying, Mapping and Cadastre for city administrative areas. The functional classes in individual years were used for further evaluation between road transport density, town development and investments into road network in the form of new by-pass, highways, etc. Following indicators showing the relation were: development of functional class areas, development of number of inhabitants, development of road transport density, and building of new decongesting roads.

VI. RESULTS

First nine cities processed in the first year of the project have brought very interesting results.

Kladno is one of processed towns situated 30 km north west from Prague. The town consists both of a core, and associated parts. The city was an industrial city in the communist period of the republic. The industrial production has been extremely declining since 1990 and most inhabitants are employed in Prague at present. Figure 1 shows spatial changes of four functional classes between 1969 and 2009 mapped by the above mentioned method. The dashed line (Figure 1 a) determines the core area as an administrative city boundary at the end of the 70-ies in the 20th century. The solid line delineates the administrative city boundary since 80-ies of the previous century which has not changed. The red color patches are residential areas in 1969. The green patches are residential areas built between 1969 and 2008.



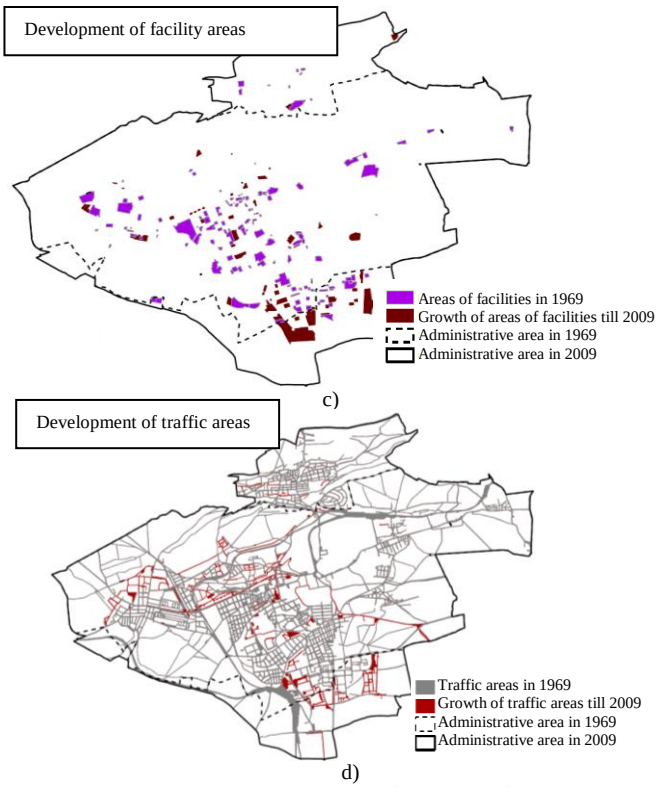


Figure 1a, b, c, d. Development of the Kladno functional classes for the 1969 – 2009 period

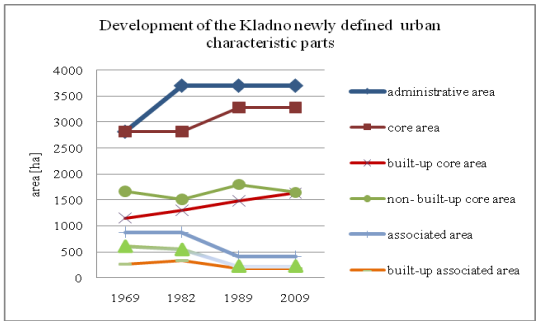


Figure 2. Example of spatial development of Kladno during last 40 years shown in administrative, core, associated, built-up and non-built-up areas

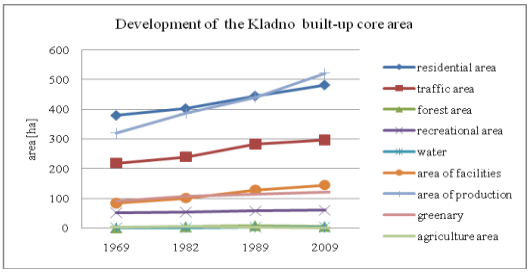


Figure 3. Example of spatial development of Kladno during last 40 years shown in land use classes

Development of production areas is shown on Figure 1b, facility areas on Figure 1c), and traffic areas on Figure 1d.

Town development is further presented on two graphs – Figure 2 and Figure 3. Figure 2 shows administrative, core and associated areas with their built-up and non-built-up area. Figure 3 represents built-up core area. Looking at the statistical evaluation presented in Figure 3, we can see that it was the area of production whose growth was the steepest in the core part. Comparing residential parts development, we can find that it covers larger areas with a steeper increase of size than those of traffic ones within the core region during last 20 years. However, there is a new highway passing the town in 5 km distance enabling the town to be used in prevailing part as a terminal location for the road traffic and not as a passing through town location. The town has not yield larger areas for recreation, leisure time, sport, etc., during last 40 years (Figure 3).

The administration area has not changed since 1982. The core area changed - unlike most towns - between 1982 and 1989. The residential area and area of production cover a similar part of the built-up area, however, the growth of the production area is steeper. Non built-up areas are in a prevailing part the forested ones. The associated areas are in most cases formed by an agricultural and forest land, however, their sizes declineafter the 1989 political change . The spatial changes are described in the coincidence table (Table 1). Each row shows the size of an individual class and its transformation between 1950 and 2008. Each column comprises original areas forming the present size of an individual class. Areas without changes are highlighted in diagonal cells of the table.Comparing of the last row (Sum 2008) and last column (Sum 1950) allows to find encrease and decrease of class areas.

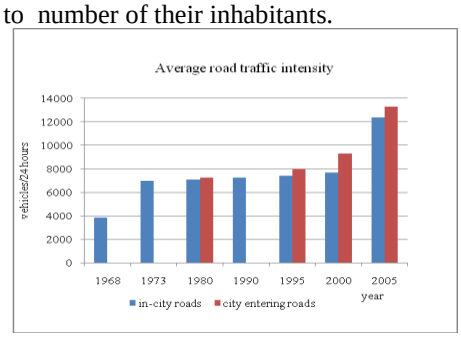
The road traffic intensity was checked both on local and higher level class roads. Both road types express a growth, however, mutually uncomparable. The slope of the growth is lower after 1995 when a new pass-by highway was built (Figure 4a). This phenomenon is presented as an impact of the highway construction out of the city on traffic intensity of the individual functional land use classes on Figure 4b.

 ISBN: 978-1-61208-118-2

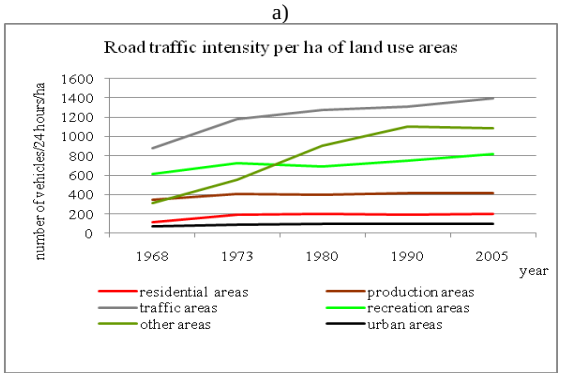
TABLE 1. THE COINCIDENCE TABLE SHOWS CHANGES BETWEEN 1950 AND 2008. RESIDENTIAL AREAS HAVE NOT CHANGED ON 207 HECTARES AND 25 % OF RESIDENTIAL AREAS HAS TRANSFORMED TO THE TRAFFIC, OTHER, PUBLIC, FACILITY, PRODUCTION, GREEN, AND AGRICULTURE AREAS (SEE THE RESIDENTIAL ROW). ON THE CONTRARY, THE PRESENT STATE ÓF THE RESIDENTIAL AREA IS NEWLY (AFTER 1950) FORMED BY TRAFFIC, OTHER, PUBLIC, GREEN, AGRICULTURE AREAS AND ENLARGED ON 130 % OF THE ORIGINAL SIZE (363 HA) (EXAMPLE OF THE CITY OF MĚLNÍK)

	Land Use Classes											
Land Use Classes	*Residential*	*Traffic*	*Forest*	*Other*	*Recreation*	*Public*	*Water*	*Facility*	*Production*	*Green*	*Agriculture*	*Sum 1950*
residential	207,47	5,56	0,00	13,06	0,00	11,79	0,00	8,32	16,94	9,98	4,01	277,14
traffic	6,24	34,55	0,95	0,34	0,16	9,40	0,00	0,21	1,83	2,86	8,16	64,97
forest	0,00	0,13	58,72	0,00	0,41	0,13	0,00	0,00	0,00	3,35	0,05	63,25
other	2,75	0,17	0,00	5,81	0,00	1,28	0,00	0,06	3,72	1,21	1,33	16,33
recreation	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
public	9,79	9,76	0,19	2,91	0,03	24,55	0,00	1,09	4,40	10,02	11,18	73,92
water	0,00	0,05	0,00	0,00	0,00	0,27	64,09	0,00	2,39	0,96	0,34	68,10
facility	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
production	0,68	1,34	0,00	3,61	0,00	2,87	1,50	0,00	66,55	3,54	0,00	80,09
green	14,09	5,47	0,00	6,94	1,63	10,40	0,00	3,27	8,66	123,28	14,66	188,88
agriculture	122,17	25,98	0,88	48,32	17,80	57,45	0,01	16,04	102,99	176,20	1082,75	1659,68
sum 2008	363,19	83,01	60,74	80,99	20,03	118,14	65,60	28,99	207,48	331,40	1122,48	

Investments into highway and by-pass road constructions can be easily recognized from two graphs in Figure 5. Ten towns with the highest number of vehicles per 24 hours entering and leavig each town were selected and compared to number of their inhabitants.



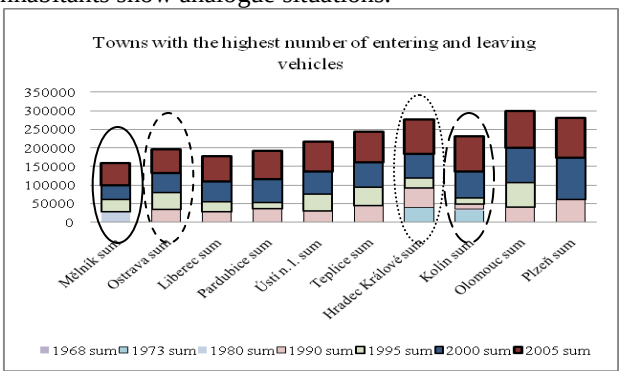
a)



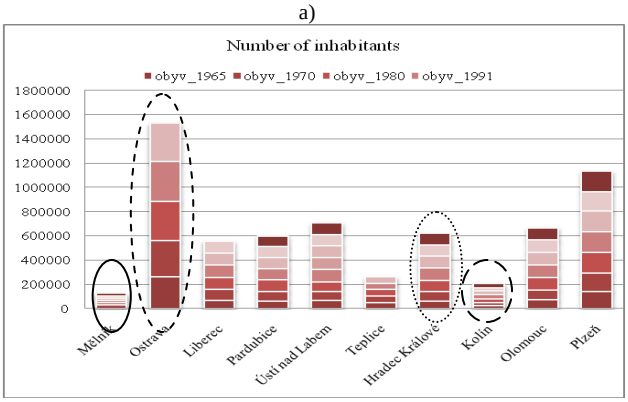
b)

Figure 4. Sum of all measured segments on higher level roads used also for passing traffic and on local roads (a). Traffic intensity calculated as a ratio of all vehicles per 24 hours and size of functional areas (b)

The important influence of by-pass roads can be found on Figure 5. The city of Mělník has a very low number and growth of inhabitants in last 40 years if compared to Ostrava, e.g.; however, numbers of measured vehicles leaving and coming to both cities are similar. Mělník does not have any by-pass road and is situated on the direction among Prague and other important Czech cities. Analyzing Kolín and Hradec Králové, their traffic intensity and number of inhabitants show analogue situations.



a)



b)

Figure 5 a, b. Comparing of the traffic intensity (a) since 1968 to 2005 and number of inhabitants (b) in similar periods (till 1991)

VII. CONCLUSIONS

The project methodology is based on a multicorrelational processing using statistical (social, economical, geographical and natural) data of cities and spatial land use development and changes. There are more than six hundred economical, social and other statistical indicators whose mutual relation are prepared to be analyzed. The spatial development visualization can be usefully presented by coincidence table.

Relations between town development and road traffic density showed interesting dependences. The traffic intensity changes cannot be generalized for a town as a unit. There are serious temporal changes within each town. These changes are caused by newly built commercial areas where this growth is incomparable to any other locations in the city, by new roads passing out of cities, e.g. The road traffic is also an important indicator of the economic inhabitant level – the traffic increase of personal and heavy- duty vehicles intensity on one side and decrease of motorcycle intensity and number of inhabitants on the other side since 70-ies is a proof of the higher economical power of the city inhabitants which was found at all fifty analyzed cities.

Results of any urban planning are always long lasting phenomena influencing the society. The spatial land use changes in 50 various cities will yield a rich source of data for the evaluation. The road traffic intensity is information on the air pollution, data on life expectancy are an issue concerning a social situation and health care, etc. The project results should offer a set of usable tools = indicators for urban planners and their further urban planning to achieve sustainable development of cities in the Czech Republic.

The paper presents a very small analysis performed for functional classes and road traffic intensity. Spatial changes and their relation to the traffic evolution have already brought a great deal of information which will be processed in a form of one of indicators. The future research is focused on determining a list and sequence of indicators for the sustainable development of cities.

ACKNOWLEDGMENT

The paper is financed by two projects: Modeling of urban areas to lower negative influences of human activities project of the Ministry of Education (OC1011) of the Czech Republic and the Management of sustainable development of life cycle of constructions, civil engineering firms and regions of the Ministry of Education project (VZ 05 CEZ MSM 6840770006) of the Czech Republic.

REFERENCES

[1] Global City Indicators Program, http://www.cityindicators.org/; cit. 30. 12. 2010.

[2] Brugmann, J., "Is there a method in our measurement? The use of indicators in local sustainable development planning" , Local Environment, vol. 2, issue 1, February 1997, pp. 59 – 72.

[3] Howell, E., Pettit, K.L.S., Ormond, B.A., and Kingsley, G.T., "Using the National Neighborhood Indicators Project to Improve Public Health, Journal of Public Health Management and Practice", vol. 9, May/June 2003, pp. 235-242.

[4] Kingsley, G.T., Kathryn, L.S., and Pettit, K.L.S., "Neighborhood Information Systems: We Need a Broader Effort to Build Local Capacity", Metropolitan Housing and Communities Policy Newsletter, October 2004.

[5] Lillesand, T.M., Kiefer, R.W., and Chipman, J.W., "Remote Sensing And Image Interpretation", 5th Ed., Wiley, 2010.

[6] Paola, J.D.; Schowengerdt, R.A., "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification", Geoscience and Remote Sensing, IEEE Transactions on, vol. 33, issue 4, July 1995, pp. 981 – 996.

[7] United Nations Human Settlements Programme, UN-HABITAT, Kenya, http://ww2.unhabitat.org/programmes/guo/urban_indicators.asp, 2003, cit 30. 12. 2010.

[8] Vepřek, K. et al, "Analysis of 100 years urban development of Hradec – Pardubice regional agglomeration focused on detection of general tendencies and regularity" - research project VÚP No16-521-503, Terplan Praha, 1983.

 ISBN: 978-1-61208-118-2