AMBIENT 2015

The Fifth International Conference on Ambient Computing, Applications, Services and Technologies

ISBN: 978-1-61208-421-3

July 19 - 24, 2015

Nice, France

AMBIENT 2015 Editors

Maarten Weyn, University of Antwerp, Belgium

AMBIENT 2015

Forward

The Fifth International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT 2015), held between July 19-24, 2015 in Nice, France, was devoted to a global view on ambient computing, services, applications, technologies and their integration.

On the way to a full digital society, ambient, sentient and ubiquitous paradigms lead the torch. There is a need for behavioral changes for users to understand, accept, handle, and feel helped within the surrounding digital environments. Ambient comes as a digital storm bringing new facets of computing, services and applications. Smart phones and sentient offices, wearable devices, domotics, and ambient interfaces are only a few of such personalized aspects. The advent of social and mobile networks along with context-driven tracking and localization paved the way for ambient assisted living, intelligent homes, social games, and telemedicine.

The conference provided a forum where researchers were able to present recent research results and new research problems and directions related to them. We welcomed technical papers presenting research and practical results, position papers addressing the pros and cons of specific proposals, such as those being discussed in the standard forums or in industry consortiums, survey papers addressing the key problems and solutions on any of the above topics, short papers on work in progress, and panel proposals.

The conference had the following tracks:

- Ambient computing environments, sensors and hardware
- Ambient devices, applications and systems
- Ambient computing and modeling

Similar to previous editions, this event attracted excellent contributions and active participation from all over the world. We were very pleased to receive top quality contributions.

We take here the opportunity to warmly thank all the members of the AMBIENT 2015 technical program committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to AMBIENT 2015. 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 also gratefully thank the members of the AMBIENT 2015 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that AMBIENT 2015 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of Ambient Computing, Applications, Services and Technologies. We also hope that Nice, France, provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

AMBIENT 2015 Chairs

AMBIENT Advisory Chairs

Maarten Weyn, University of Antwerp, Belgium
Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (ROC)
Naoki Fukuta, Shizuoka University, Japan
Andrzej Skowron, Warsaw University, Poland
Stephan Hengstler, MeshEye Consulting, USA
Andreas Schrader, University of Lübeck, Germany
Jean Vareille, Université de Bretagne Occidentale, France
Peter Roelofsma, VU - Amsterdam, The Netherlands
Tomasz Rutkowski, University of Tsukuba, Japan
Keith V. Nesbitt, University of Newcastle, Australia
Gunver Majgaard, University of Southern Denmark, Denmark
Thomas Grill, University of Salzburg, Austria
Sven Hartmann, Clausthal University of Technology, Germany
Daniel Berckmans, Katholieke Universiteit Leuven, Belgium
Eugenio Aguirre Molina, Universidad de Granada, Spain
Jean-Paul Sansonnet, LIMSI, University of Paris-Sud 11, France
José Bravo, Castilla-La Mancha University, Spain
Shusaku Nomura, Nagaoka University of Technology, Japan

AMBIENT 2015 Industry/Research Liaison Chairs

Rémi Emonet, Idiap Research Institute, Switzerland
Mieczyslaw A. Klopotek, Polish Academy of Sciences - Warszawa, Poland
Gerhard Nussbaum, Competence Network Information Technology to Support the Integration of People with Disabilities, Austria
Alexander Paar, TWT GmbH Science and Innovation, Germany
Patrick Sayd, CEA SACLAY - NANO INNOV - Gif Sur Yvette, France
Ingo Zinnikus, DFKI GmbH - Saarbrücken, Germany
Jennifer Stein, University of Southern California, USA
Florian Scharf, Institute for Multimedia and Interactive Systems, Germany

AMBIENT Special Area Chairs

Agents
Adina Magda Florea, University "Politehnica" of Bucharest, Romania

Sensing
Sandra Sendra Compte, Polytechnic University of Valencia, Spain
Context-awareness
Sylvain Giroux, Université de Sherbrooke, Canada

AMBIENT Publicity Chairs

Nancy Diniz, Rensselaer Polytechnic Institute, USA
Tomas Zizka, technical University of Liberec, Czech Republic

AMBIENT 2015

Committee

AMBIENT Advisory Chairs

Maarten Weyn, University of Antwerp, Belgium
Yuh-Jong Hu, National Chengchi University-Taipei, Taiwan (ROC)
Naoki Fukuta, Shizuoka University, Japan
Andrzej Skowron, Warsaw University, Poland
Stephan Hengstler, MeshEye Consulting, USA
Andreas Schrader, University of Lübeck, Germany
Jean Vareille, Université de Bretagne Occidentale, France
Peter Roelofsma, VU - Amsterdam, The Netherlands
Tomasz Rutkowski, University of Tsukuba, Japan
Keith V. Nesbitt, University of Newcastle, Australia
Gunver Majgaard, University of Southern Denmark, Denmark
Thomas Grill, University of Salzburg, Austria
Sven Hartmann, Clausthal University of Technology, Germany
Daniel Berckmans, Katholieke Universiteit Leuven, Belgium
Eugenio Aguirre Molina, Universidad de Granada, Spain
Jean-Paul Sansonnet, LIMSI, University of Paris-Sud 11, France
José Bravo, Castilla-La Mancha University, Spain
Shusaku Nomura, Nagaoka University of Technology, Japan

AMBIENT 2015 Industry/Research Liaison Chairs

Rémi Emonet, Idiap Research Institute, Switzerland
Mieczyslaw A. Klopotek, Polish Academy of Sciences - Warszawa, Poland
Gerhard Nussbaum, Competence Network Information Technology to Support the Integration of People with Disabilities, Austria
Alexander Paar, TWT GmbH Science and Innovation, Germany
Patrick Sayd, CEA SACLAY - NANO INNOV - Gif Sur Yvette, France
Ingo Zinnikus, DFKI GmbH - Saarbrücken, Germany
Jennifer Stein, University of Southern California, USA
Florian Scharf, Institute for Multimedia and Interactive Systems, Germany

AMBIENT Special Area Chairs

Agents
Adina Magda Florea, University "Politehnica" of Bucharest, Romania

Sensing
Sandra Sendra Compte, Polytechnic University of Valencia, Spain

Context-awareness
Sylvain Giroux, Université de Sherbrooke, Canada

AMBIENT Publicity Chairs

Nancy Diniz, Rensselaer Polytechnic Institute, USA
Tomas Zizka, technical University of Liberec, Czech Republic

AMBIENT 2015 Technical Program Committee

Sachin Kumar Agrawal, Samsung - Noida (Delhi NCR), India
Eugenio Aguirre Molina, Universidad de Granada, Spain
Cristina Alcaraz, University of Malaga, Spain
Hani M. Alzaid, King Abdulaziz City for Science and Technology, Saudi Arabia
Cecilio Angulo, Universitat Politècnica de Catalunya, Spain
Eduard Babulak, Sungkyunkwan University, Korea
Costin Badica, University of Craiova, Romania
Youakim Badr, INSA - Lyon, France
Maxim Bakaev, Novosibirsk State Technical University, Russia
Flavien Balbo, University of Paris Dauphine, France
Stephen Barrass, University of Canberra, Australia
Jean-Paul Barthès, Université de Technologie de Compiègne, France
Ben Bellekens, University of Antwerp, Belgium
Cecile Belleudy, University Of Nice - Sophia Antipolis LEAT laboratory, France
Orlando Belo, University of Mihno, Portugal
César Benavente, Universidad Politécnica de Madrid, Spain
Daniel Berckmans, Katholieke Universiteit Leuven, Belgium
Gerald Bieber, Fraunhofer IGD, Germany
Daniel Biella, University of Duisburg-Essen, Germany
Stephan Böhm, Hochschule RheinMain, Germany
Lars Braubach, University of Hamburg, Germany
José Bravo, MAmI Research Lab, Spain
Philip J Breedon, Nottingham Trent University, UK
Ramon F. Brena Pinero, Tecnológico de Monterrey, Mexico
Valerie Camps, Universite Paul Sabatier - Toulouse, France
Juan-Carlos Cano, Universitat Politecnica de Valencia, Spain
Carlos Carrascosa, Universidad Politécnica de Valencia, Spain
John-Jules Charles Meyer, Utrecht University, The Netherlands
Omar Cheikhrouhou, University of Sfax, Tunisia
Albert M. K. Cheng, University of Houston, USA
Cezar Collazos, Universidad del Cauca, Columbia

Carlos Fernando Crispim-Junior, INRIA Sophia Antipolis, France
Manuel Cruz, Immersion Corp, Canada
Giuseppe De Pietro, High Performance Computing and Networking Institute - National Research Council, Italy
Jianguo Ding, University of Skövde, Sweden
Nancy Diniz, Rensselaer Polytechnic Institute, USA
Jeroen Doggen, Artesis Plantijn University College Antwerp, Belgium
Wael Mohamed Elmedany, University of Bahrain, Bahrain
Abdelkarim Erradi, Qatar University, Qatar
Larbi Esmahi, Athabasca University, Canada
Andras Farago, University of Texas at Dallas, USA
Juan Manuel Fernández, Barcelona Digital Technology Centre, Spain
Scott S. Fisher, USC School of Cinematic Arts, USA
Adina Magda Florea, University "Politehnica" of Bucharest, Romania
Jesús Fontecha Diezma, Universidad de Castilla-La Mancha, Spain
Peter Forbrig, University of Rostock, Germany
Peter Froehlich, Telecommunications Research Center Vienna, Austria
Francesco Furfari, ISTI-CNR, Italy
Kaori Fujinami, Tokyo University of Agriculture and Technology, Japan
Naoki Fukuta, Shizuoka University, Japan
Mathieu Gallissot, Université Joseph Fourier - Grenoble, France
Matjaz Gams, Jožef Stefan Institute - Ljubljana, Slovenia
Marisol García Valls, Universidad Carlos III de Madrid, Spain
Amjad Gawanmeh, Khalifa University of Science, U.A.E.
Hassan Ghasemzadeh, Washington State University, USA
Michele Girolami, ISTI-CNR / University of Pisa, Italy
Sylvain Giroux, Université de Sherbrooke, Canada
Thomas Grill, University of Salzburg, Austria
Hans W. Guesgen, Massey University, New Zeeland
Maki K. Habib, American University in Cairo, Egypt
Lynne Hall, University of Sunderland, UK
Sven Hartmann, Clausthal University of Technology, Germany
Fumio Hattori, Ritsumeikan University - Kusatsu, Japan
Peter Hellinckx, KdG University College, Belgium
Ramon Hervas, Castilla - La Mancha University, Spain
Richard Hoadley, Anglia Ruskin University - Cambridge, UK
Michael Hobbs, Deakin University - Geelong, Australia
Patrick Horain, Institut Mines-Telecom / Telecom SudParis, France
Marc-Philippe Huget, University of Savoie, France
Achilles Kameas, Hellenic Open University, Greece
Martin Kampel, Vienna University of Technology, Austria
Anthony Karageorgos, Technological Educational Institute of Larissa - Karditsa, Greece
Ridha Khedri, McMaster University, Canada
Mehdi Khouja, University of Gabes, Tunisia

Yacine Sam, Université François-Rabelais Tours, France
Nayat Sanchez Pi, Universidade Federal Fluminense, Brazil
André C. Santos, Microsoft Language Development Center, Portugal
Patrick Sayd, CEA SACLAY - NANO INNOV - Gif Sur Yvette, France
Andreas Schrader, University of Lübeck, Germany
Boris Schauerte, Karlsruhe Institute of Technology, Germany
Thomas Schlegel, Technische Universität Dresden, Germany
Shishir Shah, University of Houston, USA
Charles G. Sheridan, Intel, Ireland
Andrzej Skowron, Warsaw University, Poland
Jennifer Stein, University of Southern California, U.S.A.
Weilian Su, Naval Postgraduate School, USA
Dante I. Tapia, University of Salamanca, Spain
Teck Hou Teng, Nanyang Technological University, Singapore
Takao Terano, Tokyo Institute of Technology, Japan
Sergi Torrellas, Barcelona Digital Technology Centre, Spain
Markus Ullmann, Bundesamt für Sicherheit in der Informationstechnik (BSI) | University of Applied Sciences Bonn-Rhine-Sieg, Germany
Rainer Unland, University of Duisburg-Essen, Germany
Jean Vareille, Université de Bretagne Occidentale, France
Eloisa Vargiu, Barcelona Digital Technology Center, Spain
George Vouros, University of the Aegean - Samos, Greece
Thomas Waas, Ostbayerische Technische Hochschule (OTH) Regensburg, Germany
Benjamin Weyers, RWTH Aachen University, Germany
Markus Wiemker, RWTH Aachen University, Germany
Lawrence W.C. Wong, National University of Singapore, Singapore
Koichi Yamada, Nagaoka University of Technology, Japan
Kin Choong Yow, Gwangju Institute of Science and Technology, South Korea
Chang Wu Yu, Chung Hua University, Taiwan
Zhiyong Yu, Fuzhou University, China
Olivier Zendra, INRIA Nancy, France
Ingo Zinnikus, DFKI GmbH - Saarbrücken, Germany

Table of Contents

Effective Mission Management through Service-aware Streaming Infrastructure 1
Marc Roelands, Wolfgang Van Raemdonck, Stephane Menoret, Laurent Navarro, Ana Bildea, and Sebastien Creiche

Monitoring Abnormal Behavior of Hospital Patients Using RGB+D Sensors 10
Seung Chae and Kin Choong Yow

Flexible Management of Ambient Assisted Living Environments by Ontology-Based Configuration 16
Tom Zentek, Jurgen Bock, and Asarnusch Rashid

Geolocation Architecture for Combining Multiple Technologies 21
Jean-Baptiste Drucker, Tony Ducrocq, Colombe Herault, Guillaume Lefebvre, and Eric Ptak

An "Internet of Things" Vision of the Flood Monitoring Problem 26
Massimo Ancona, Andrea Dellacasa, Giorgio Delzanno, Andrea La Camera, and Ivano Rellini

A Smartwatch-Based System for Audio-Based Monitoring of Dietary Habits 30
Haik Kalantarian and Majid Sarrafzadeh

k-Nearest-Neighbour based Numerical Hand Posture Recognition using a Smart Textile Glove 36
Cai Li, Katharina Bredies, Anja Lund, Vincent Nierstrasz, Paul Hemeren, and Dan Hogberg

Component Templates and Service Applications Specifications to Control Dynamic Adaptive System Configurations 42
Holger Klus, Andreas Rausch, and Dirk Herrling

Touch User Interface Evaluation for E-book Systems: A Usability Experiment on Zinio and MagV 52
Yang-Cheng Lin and Chung-Hsing Yeh

Improvement on K-means and TSP Based Mobility Protocol of Wireless Sensor Network 56
Afef Ghabri, Monia Bellalouna, and Walid Khaznaji

New Geometric Geolocalization Method for Tracking Patients in Medical Environments 65
Afef Ghabri, Monia Bellalouna, and Leila Horchani

A Framework for Supporting Natural Interaction with Printed Matter in Ambient Intelligence Environments 72
George Margetis, Margherita Antona, and Constantine Stephanidis

Multi-Target Data Association in Binary Sensor Networks 79
Sebastian Matthias Muller and Andreas Hein

Effective Mission Management through Service-aware Streaming Infrastructure

Marc Roelands,
Wolfgang Van Raemdonck
Bell Labs, Alcatel-Lucent
Antwerp, Belgium
Email: {marc.roelands,
wolfgang.van_raemdonck}
@alcatel-lucent.com

Stéphane Ménoret,
Laurent Navarro
Communication & Security, Services
Thales Group
Paris, France
Email: {stephane.menoret,
laurent.navarro}@thalesgroup.com

Ana Bildea,
Sébastien Creiche
Arago Systems
Nice, France
Email: {ana.bildea,
sebastien.creiche}
@aragosystems.com

***Abstract*— In this paper, we introduce an innovative, service platform-based approach to dynamic, world model-derived stream prioritization and selection, going beyond today's practice in urban security solutions. By validating the approach in a realistic urban simulation and resource-constrained wireless sensor network context, we demonstrate a significant improvement in situation awareness and effective resource management for security mission operations.**

Keywords-urban security; surveillance mission management; dynamic stream control; service platform; real world simulation; wireless sensor network.

I. Introduction

With the trend towards *Internet of Things* and the emergence of ubiquitous wireless connectivity in general, the capability to stream information, live and on demand, is pervading our society ever more, also extending the potential for urban security solutions. States and governments, ministries of defense and homeland security agencies, but by delegation also critical infrastructure operators or large private security agencies, aim to protect citizens and public infrastructure, and aim to monitor and act upon emergencies with various damage-mitigating actions, on a local up to internationally collaborative scale. With those aims, they invest in dedicated infrastructure roll-outs, potentially leveraging public / privately owned civilian infrastructure as well.

From a value proposition canvas perspective [1], the typical *pains* that organizations in charge of security face are: the unpredictability of each new security situation, the practical heterogeneity of information acquisition systems to be integrated, the unfeasibility of manual browsing through the available abundance of information streams (especially during crisis situations), and personnel budget scarcity (leading to cognitive overload of teams budgeted too small – as one operational person traditionally is assumed capable of handling around ten simultaneous video feeds, surveillance in large cities with thousands of cameras implies quite an extensive staffing).

Consequently, organizations in charge of security are seeking solutions that provide: faster incident-to-safety return response time during missions, flexibility to define before each mission the tool-based support needed, proactive situation awareness support, ranking information according to its relevance to the (dynamic) situation, and autonomous system behavior with respect to system overload handling in (often also mission-critical) high-load operational phases.

We see such requirements confirmed in examples such as the surveillance system in Mexico City [2]. Since 2010, more than 8000 video cameras, gunshot detectors, and license plate recognition systems, including even unmanned aerial vehicles (UAV), have been deployed in this case, requiring an extensive infrastructure with several C2 (Control and Command) and higher level C4I (Command, Control, Communications, Computers, and Intelligence) centers, and a force of 3000 specialized police agents operating the centers.

Security missions, such as high profile events and VIP protection (VIP: Very Important Person, e.g., a nation's president), are not yet fully managed by today's surveillance solutions, as they typically involve temporary, ad-hoc rolled-out security mission field infrastructure and services. From a security assessment point view [3], the main critical asset to protect then becomes the VIP, being subject to various lethal threads, including potential terrorist attacks in visited public areas. VIP protection, the live screening, monitoring and tracking of the VIP, as well as of suspected-malicious people, is complicated by a number of typical mission constraints. The area to be covered by a mission may be wide, freely accessible to the public and overcrowded, e.g., an exhibition centre with both indoor and outdoor sections to which a VIP visit was publicly announced. It may at the same time however be desirable to keep the actual protective measures hidden from the general public. Moreover, each mission follows a unique scenario, implying that history or routine from other missions often cannot be reused straightforwardly. A pre-assessment of the situation is performed before each mission, and a corresponding (multi path option) evacuation plan is typically prepared.

Next to the operational asset protection preparations and related risks, the rolled-out technical infrastructure itself also may have its vulnerabilities. Mission management often critically depends on a wireless network that can be affected by jamming, communication interception and replay. Therefore, mesh network technology, with link redundancy and multiple simultaneously used radio technologies are considered. In security missions where video streaming is

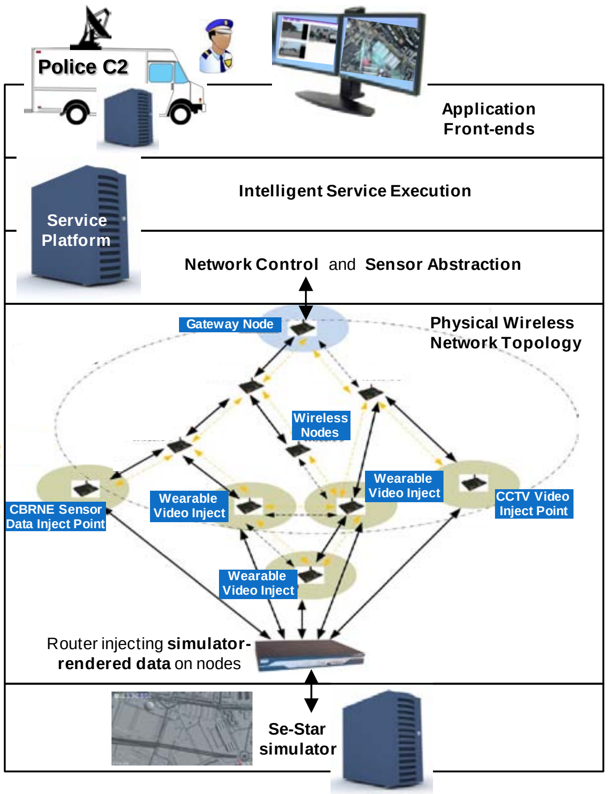


Figure 1. Overview of the experimental setup.

needed, the solutions should especially be able to cope with the limited bandwidth and battery autonomy of the ad-hoc rolled-out wireless network nodes.

In the next section we propose an innovative solution providing a considerable improvement over the existing solutions in use today for security mission management. Section III elaborates on the intelligence we established on the service platform, and Section IV makes a first qualitative and quantitative evaluation of the solution.

II. Mission Management Solution

Considering the mentioned security mission management challenges, we designed a solution allowing:

- *flexible addition of any required sensors*, be it ad-hoc for a mission or accessed via permanently installed, possibly privately owned infrastructure;
- *flexible addition / customization of mission services* (at a programmable service priority) that can *easily be launched during missions*;
- *situation awareness platform support* as relevant to the services requested during missions, so as to *lower cognitive overload* for the security team; and
- *autonomous prioritization of the (video) stream load* on the wireless network (links, nodes and devices), as dynamically dependent on the needs of the requested services (and so also leveraging the same platform situation awareness).

We distinguish following solution infrastructure layers:

- the physical (typically geographically spread) **devices** sensing the real world - our experiments include the simulation of, and content rendering for wearable and CCTV (Closed-Circuit Television) cameras, GPS (Global Positioning System) localization, and specialized toxic particles-detecting (CBRNE, Chemical, Biological, Radiological, Nuclear and Explosives) sensors,
- the **network** connecting the sensor range to a central service platform - our experiments use a prototype dual radio sensor board network,
- a **network control** and **sensor abstraction** platform level - our experiments use SDN-like (SDN: Software-Defined Networking) network control [4] and OGC-compliant (OGC: Open Geospatial Consortium) sensor abstraction [5],
- an **intelligent service execution** platform level - our experiments use a single-machine instance of the generic service platform running at a mobile C2 centre,
- a set of **application front-ends**, providing a user interface to the C2 team for video surveillance and CBRNE sensor monitoring services.

As a solution blueprint, applicable more broadly than in the security mission cases of our experiments, the design assumes a *horizontal, generically reusable* intelligent service platform. Key to the intelligence of this platform is that it leverages (service-independent) **world models** for enhanced situation awareness and prediction. While our experiments show that elementary approximations of such models can already be effective for the optimizations the platform provides, it is absolutely crucial to the *validation* of the solution to be able to embed it in a context of realistic and live real-world data streams. Particularly, as security missions cannot be easily 'rehearsed' in the real world, we validated the solution by means of the **sector-professional human behavior simulator SE-Star**, which we instrumented to provide the control interfaces and to render the (video and sensor) streaming content as would be available from the actual physical devices in actual field operation scenarios. (Such simulation is also used commercially today for system validation, field force training and large-scale exercises.) During actual missions, cameras and sensors (worn by policemen, ad-hoc fixed to walls, or as groups of fixed CCTV infrastructure) would each be connected to a sensor network node, with a *gateway* node eventually connecting the ad-hoc mesh network to the service platform in the C2 centre. Figure 1 summarizes the experimental setup, allowing to realistically validate the solution's effectiveness for real-world scenarios, without the actual real geo-physical deployment and action.

We consider a **guidance and evacuation scenario during a VIP visit** at a large exhibition center in Paris, considering also a potential toxic bomb threat, as the validation case. We use a realistic 3D mesh model of the exhibition center in SE-Star. With the SE-Star human behavior modeling (e.g., VIP-following or panic-motivated) and environment features (e.g., toxic gas cloud dispersion and impact), even a terrorist attack can be realistically simulated in the scenario.

 ISBN: 978-1-61208-421-3

From the application front-ends, that in practice would be installed in a truck near the mission scene, the C2 staff coordinates the protection team in the exhibition center, and should thus achieve effective live situation awareness concerning relevant events in the mission scene. For that purpose, an actual Wireless Sensor Network (WSN) is included in the experimental setup, as would also be rolled out in practice. In the scenario, the ad-hoc network is considered to be connected to the exhibition center's CCTV network. As such, about thirty cameras and a similar amount of chemical sensors, a GPS sensor worn by the VIP, and several cameras worn by policemen are connected to the physical network from Se-Star, at the corresponding network node boards. Their data streaming can be controlled from the service platform, as such dynamically injecting the data in the physical network.

Actual services made available to C2 staff in the service platform for activation during the mission (in line with what would be prepared during an actual mission planning phase) are:

- a *scene overview* service, providing the C2 staff with a general overview of the area they plan the VIP to be visiting,
- a *person monitoring and guidance* service, allowing the C2 staff to monitor the VIP on a planned visit trajectory, and command and guide evacuation when needed,
- a *crowd monitoring* service, allowing the C2 staff to detect and track crowds dynamically occurring in the scene (potentially hindering VIP evacuation), and
- a *toxic gas cloud monitoring* service, allowing the C2 staff to detect and observe the live impact of a chemical/bomb incident (complemented with a front-end for detailed CBRNE sensor readings) (again impacting VIP evacuation).

The next subsections zoom in on our WSN prototype and SE-Star.

A. *Wirless Sensor Network prototype*

We composed our WSN of extendible board prototypes as shown in Figure 2, each of which has a high (10 Mbps) and low (250 kbps) data rate unit. They support required higher level protocols such as CoAP (Constrained Application Protocol) and OGC SOS (Sensor Observation Service) [5]. (OGC SOS is used as a southbound interface to the service platform, as a standards-based way to interact with wireless sensors.)

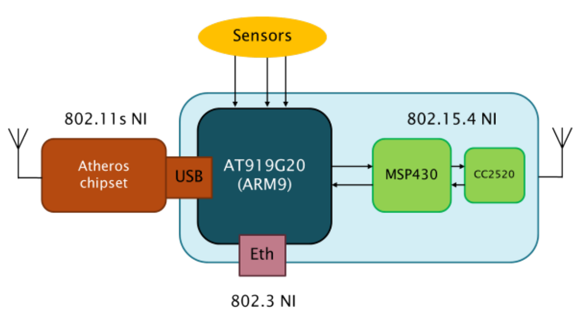


Figure 2. Network node block diagram

The **high data rate hardware unit** runs Linux on a 32-bit ARM processor with 64 MB SRAM (and extendible Flash memory), and has a IEEE 802.11b/g/n USB dongle for the video stream transport (and wired Ethernet for simulated stream injection). We use a IEEE 802.11s driver and Hybrid Wireless Mesh Protocol (HWMP) for routing the video streams in the typical Multi-Point-to-Point situation of the class of use cases at hand. Videos are streamed using RTP/RTSP (Real time Transport Protocol / Real Time Streaming Protocol) [6][7]. A CGI (Common Gateway Interface) web server is used for board configuration (particularly, flow configuration from the service platform).

The **low data rate / low power hardware unit** runs the small footprint Contiki operating system supporting IEEE 802.15.4 and 6LoWPAN [6], on a TI MSP430 16-bit microcontroller with 16 KB internal RAM and 256 KB Flash, with the TI CC2520 chipset for IEEE 802.15.4 low power networking in the 2.4 GHz ISM band. The Contiki IPv6 stack offers low power standard RPL (Routing Protocol for Low-Power and Lossy Networks) routing [7] and CoAP [8], making it ideal for the CBRNE sensor data video control streams.

We designed for **interoperation of the two hardware units**, allowing in principle to turn to an energy-saving low power mode for signaling, configuration and low bandwidth sensor data streaming only, at times when no video needs to be routed through a particular network node. Each video stream has been measured to add 150 mW of power consumption to a network node (added to a base consumption of 750 mW, which could be lowered in idle state by means of a duty cycle mechanism).

Figure 3 shows the network topology as enforced on the boards in the lab (where network nodes were laid out on tables), corresponding to the actually depicted *geographically spread* mesh organization that would be imposed in reality. (The manual enforcement is needed in the lab setup, because, without this, the nodes would connect as a full IEEE 802.11s mesh, while in reality the geographical spreading of the nodes would prevent links to exist between geo-distant nodes.) In the figure, dashed lines represent existing radio links in the topology whereas solid lines represent a chosen set of default routes within the topology.

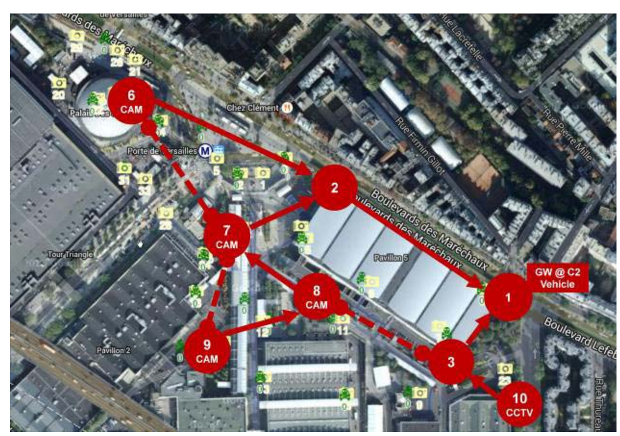


Figure 3. Enforced wireless network toplogy

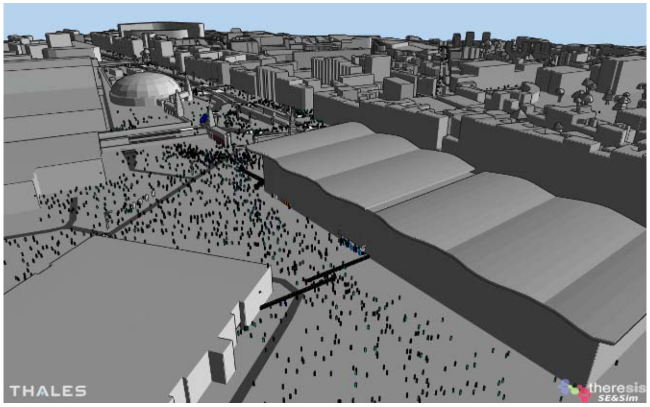


Figure 4. SE-Star urban security city simulator

Given a practical 5 Mbps driver limitation observed for the high data rate hardware unit and the known mesh multi-hop limitation (per-hop bandwidth halving in shared medium), a practical maximum of 5 video streams was observed for the actual setup (at an average 200 kbps per 'high quality' stream instance). This demonstrates that – even in later improved production-ready hardware conditions – it is of high importance to **smartly select streams dynamically**, and to use the alternative low data rate network for critical control signals.

B. SE-Star human behavior simulator

SE-Star, as illustrated by Figure 4, is a multi-agent simulator focusing on reproducing human behavior in large-scale environments [9]. It aims to provide an adaptive and modular tool for planning, decision-making and training purposes, on any kind of real or fictive scenario. To do so, it relies on a bio-inspired motivational engine, animating thousands of individual agents within real-scale critical infrastructures, interacting with each other and with objects reproducing real-life equipment. SE-Star is fully customizable, allowing the user to define the environment mesh, the agents' and objects' characteristics and behaviors, and the scenario script, in an easy way. It thus provides a reliable and realistic simulation basis for our experiments.

III. LEVERAGING REAL WORLD KNOWLEDGE FOR DYNAMIC OPTIMIZATIONS

With the experimental setup where the SE-Star virtual cameras and sensors can be controlled to inject requested live data and rendered video into the physical entry points of the network, configured in a realistically enforced network topology, the service platform can control and manage streams in a **fully realistic context**, and can be evaluated on serving the actual case-specific applications effectively. The intelligent service execution level of the service platform selects and prioritizes video stream loading of the network using **real world knowledge modeling** as a key enabling mechanism to add situation awareness and prediction in the service execution context. The mechanism allows the service platform to **dynamically select** the most critically needed and most relevant video streams for each service context, while simultaneously determining an overall platform **prioritization** for those streaming needs for the near future, as a dynamic and proactive means of network route and resource reservation.

Applicable beyond our current experiments, some platform design aspects need to be noted.

First of all, we make an explicit distinction between any *service goals* that may be requested by a user (i.e., during a mission in this case) and the real world facts expressed in the *world model*, which are pre-articulated by a *domain expert.* Service composition can be done by means of *service templates*, making referencing of world model elements from the templates straightforward. This serves as an **inherently scalable context-awareness approach**. Indeed, as if it were, context engine 'model fractions' are *woven into* the service composition during the process, inherently scaling up context processing according to service instance needs. (Sub-Section II.A. will discuss how this is achieved.) This thus avoids the well-known bottleneck seen in traditional presence servers in communication services [9] or the similarly implied need for elasticity solutions in publish-subscribe systems [10].

Using a world model in this way, in general, enables **proactive service-aware resource management** of the network and the deployed service processing, resulting e.g. in dynamic bandwidth reservation or distributed code placement, complementing reactive service-agnostic resource management autonomously by the system itself. In the case focused on in this paper, we consider de constrained WSN links as the resources to be proactively managed.

A. Service composition with knowledge weaving

The service platform thus has a knowledge base storing world model elements, describing the **behavioral constraints of particular real-world phenomena,** and a set of service templates. Figure 5 shows an example of such a phenomenon behavioral constraint as used for the current prototype use case. In this example, we express a first order prediction of *the movement of a person* for which the current position can be observed and a set of course waypoints are known (as the case with a prepared VIP visit plan). Particularly, the code in Figure 5 shows a function representing this knowledge, returning a person's 2D geo-position predicted for a time *lead_time* ahead, based on the current observed position *c_pos* (at current time *c_t*), the last observed position *l_pos* (at time *l_ts*), and a set of planned positions (*path*) from course waypoint known from the VIP visit plan. The (somewhat arbitrarily chosen) heuristic captured by the function is that the extrapolation based on the speed estimate is vector-wise corrected by averaging its direction with the direction perpendicular to the line segments of the nearest course planned waypoints (i.e. towards the planned path).

Models of similarly elementary nature are devised for other relevant phenomena in the application domain and validated in the experimental setup, most notably, occurrence and prediction of *crowds*, based on observed people density and individual speeds, and predicted *dispersion of toxic gas*


```
ppos_predict(lead_time, path, (c_ts,c_pos), (l_ts, l_pos))
    dt = l_ts - c_ts
    speed = dist/dt
    vip_dir= track.compute_ori(l_pos, c_pos)
    to_path= perpendicular_lines(path, c_pos)
    est_dir=track.normalize(vec2_sum(vip_dir,
                                     track.normalize(to_path)))
return track.latlon(vec2_sum(c_pos,
                    vec2_mul(est_dir, speed*lead_time)))
```

Figure 5. Moving person as example of a real-world phenomenon; above: planned trajectory on map; below: phenomenon world model fragment

clouds, based on above-threshold CBRNE sensor measurements and a first order circular expansion model.

Much more complex world models could be used, where available or when they can be generated. When also sufficient live observation data would be available, this could even lead to more accurate predictions. However, our experiments have shown that the used of just **coarse estimates can already be turned into considerable operational advantage**.

As an example service template, all *monitoring* type of services in our experiments use the service template shown in Figure 6. The service templates are implemented as highly parameterized directed graph descriptions of connected execution primitives. In the example, with parameters in bold text, the graph composes the essential elements of phenomenon visual tracking, pan-tilt-zoom control of cameras, selection and activation of camera streams and prediction of such for network control, and ultimately video mixing for displaying the video streams in the designated user interface area. Apart from the set of camera sources that

```
showme [icore:phenomenon]
       [icore:camera/set]
       [icore:sink]  {
[icore:phenomenon/observation/pos] -> geos_filter ->
    select_ctrl -> ptz -> axis_ctrl[icore:camera/ptz_config];
[icore:phenomenon/observation/pos] ->
    ptz[icore:camera/ptz_config];
[icore:camera/set] -> geos_filter[icore:camera/coverage] ->
    c2service -> streams -> flow_ctrl;
c2service -> mixer -> [icore:sink/video];
streams -> [icore:camera/video] -> mixer;
[icore:phenomenon/observation/prediction] -> flow_ctrl;
}  [icore:showme_service]
```

Figure 6. Example service template logic

the service should consider (in this case, all cameras available in the exhibition centre as registered for the mission), destination sink for the (raster-mixed) video streams (in this case, always a designated area in the C2 application front-end user interface, i.e., network-wise behind the gateway node of the WSN), the real-world phenomenon to be observed and status-predicted is a template parameter. When instantiating the template, it is fed with the appropriate phenomenon description, e.g. the observation and status-prediction of a person, in this case the VIP. The knowledge base holds the implications of that choice, e.g., the type of further sensor input (Global Positioning System readings of the device carried by the person), the way how to observe and track the person via a camera, and how to predict the behavior of the person, e.g., how to predict position, i.e., the example behavioral knowledge expressed in Figure 5. In this way, the **full description of the service instance graph can be derived for a *service request***, as a weaving of the applicable service template logic and the applicable phenomenon behavioral knowledge.

B. *Execution of service instances*

The system thus *instantiates* a service template upon receiving a user-issued *service request*, extracting parameter values from the request, and referencing any relevant real world phenomena behavior from the knowledge base. Upon full composition of the requested service instance (as was discussed in Section III.A.), the executable description of the service instance is **deployed and started as a data stream processing graph**.

From the template example in Figure 6, we see that all service instances produced from that template have their graph wired in the overall execution graph of Figure 7 according to the links denoted with "->" in the template. Based on their nature, some of the composing processing nodes are instantiated per service instance (e.g., mixer), while others are common to all service instances or considered infrastructure components (e.g., flow_ctrl and streams, as the *Flow Ctrl* and *Stream Mngr* blocks displayed in Figure 7, respectively). The template parameters (bold in Figure 6) are used to inject specific bindings, among which also the specific phenomenon observed, as corresponding to the issued service request. E.g., when a *person monitoring and guidance* service is requested by the C2 staff, a service instance is deployed and executed filling the template with the observation and prediction logic for *movement of a person* is used for the phenomenon parameter. This implies via further template and world model elaboration, e.g., that the function ppos_predict from Figure 6 is embedded in a processing node called path_estimation in Figure 7.

Figure 7 thus shows four such concurrent service instances resulting from service requests in the example scenario experiments. The shaded ellipsoid graph parts show where world model dependencies are inserted, thus leveraging properties of one or more relevant phenomena either for selecting or controlling cameras, or for predicting near-future observations of the phenomenon, for requesting a corresponding network provisioning. The *Stream Mngr*

 ISBN: 978-1-61208-421-3

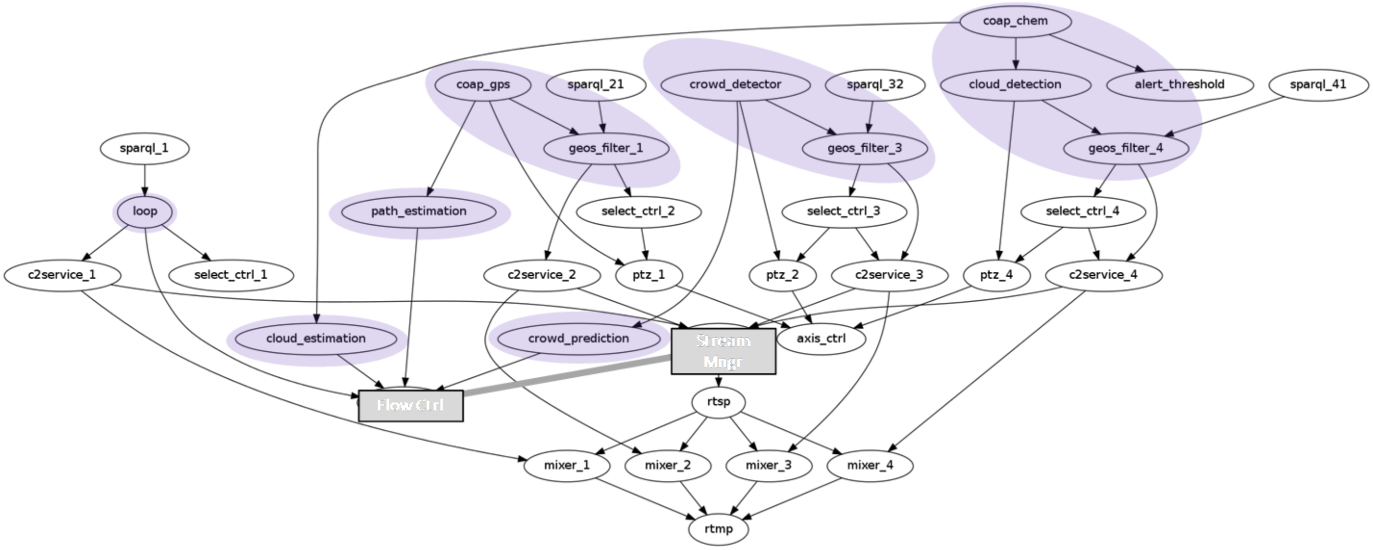


Figure 7. Example data processing graph resulting from 4 example service requests

block collects the camera selection requests and handles the actual stream control for the cameras over the network.

C. *Service-aware stream priority control*

In analogy to the SDN network control concept [4], the service platform considers the notion of *'flow'* as a programmability abstraction for the network. As part of the service platform, a **flow controller** (*Flow Ctrl* block in Figure 7), associated one-on-one to a particular network, dynamically declares which **flows** – defined by a source and destination address and required **QoS** (Quality of Service) characteristics – need to get a particular **(reserved) network priority**. When an actual **stream** is requested to be transported over the network by the **stream manager** (*Stream Mngr* block in Figure 7), after a given maximum *network reconfiguration time* $t_{maxconf}$, the stream gets assigned to a particular flow, and so is handled according to the flow's declared priority and QoS. In the experimental setup, flows correspond to particular camera to network gateway combinations and an associated bandwidth requirement, as a simplified QoS characteristic, either "High Quality" (HQ), corresponding to an appropriate stream bandwidth for a "main" video screen area of a service, or "Low Quality" (LQ), for all other video streams.

The flow controller determines overall network flow priorities and QoS level dynamically by linearly combining the relative priorities predicted for each candidate requested stream for each service instance, based on the priority of *each service instance* (as determined by the C2 staff) and the dynamic *likelihood of relevance of each candidate stream* in that service instance. The dynamic likelihood of relevance for the candidate streams is determined by the respective services' phenomenon predictors (arrows from prediction graph edges to *Flow Ctrl* block in Figure 7), which thus implies predictions need to happen $t_{maxconf}$ ahead of time, to allow for network reconfiguration.

Furthermore, the flow controller takes into account the *limited capacity of the network* (overall or for particular flows) ensuring no overloading in any upcoming time interval, thus timely and dynamically provisioning the network for only the highest-priority-ranked flows. The stream manager consequently only allows actual streams in provisioned flow paths, and **blocks non-flow-reserved stream requests** to prevent network overloading.

In the current setup, the actual **routing** for requested flows is dynamically chosen according to one out of a set of overall routing plans that have been pre-determined to be suitable for the network topology at hand. As flow requests arrive to it, the flow controller thus heuristically requests the network to keep the current, or switch to another overall routing plan. The typical mesh bandwidth division effect and the shared medium indeed essentially limit the transmission capacity of the particular network setup, rendering finer-grain routing optimization for overall bandwidth improvement less useful for the particular use case. Beyond bandwidth utilization, the heuristic routing plan switching is beneficial for **load spreading across the network**, avoiding individual wireless network node battery drain, which would degrade the network critically.

IV. EVALUATION

We conducted a range of experiments, assuming the VIP guidance and evacuation scenario, and typical C2 user interactions and user service requests. In the next subsections, we revisit the requirement assumptions, the resource use effectiveness, and the actual mission management effectiveness, as concluded from the real-world validation context for the intelligent service platform.

A. *Qualitative validation of requirements*

Using the realized prototype as a premise, we interviewed commercial domain experts. They confirmed a correct requirements focus, and recognized the operational

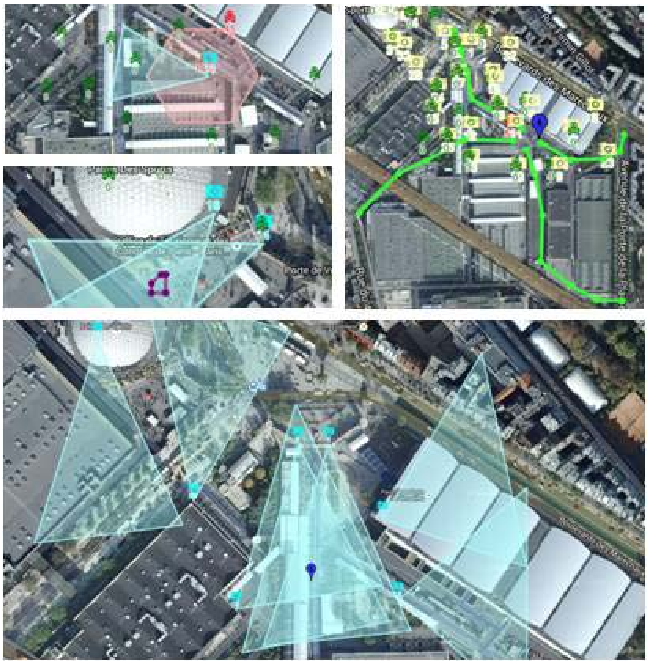


Figure 8. Snapshot impression of live geo-overlay C2 user feedback: (top left) crowd and toxic gas cloud anticipations, (top right) candidate VIP evacuation paths, (bottom) active-stream camera orientation

improvements targeted by our system as highly valuable indeed, however adding as a remark that, even when kept rudimentary, the **world model, being an evident dependency** for the system should be provided by an ‘impact-conscious’ domain expert, for ensuring reliable system behavior. As another remark, C2 staff should anyhow also be able to **verify and potentially override the decision support** stream selection.

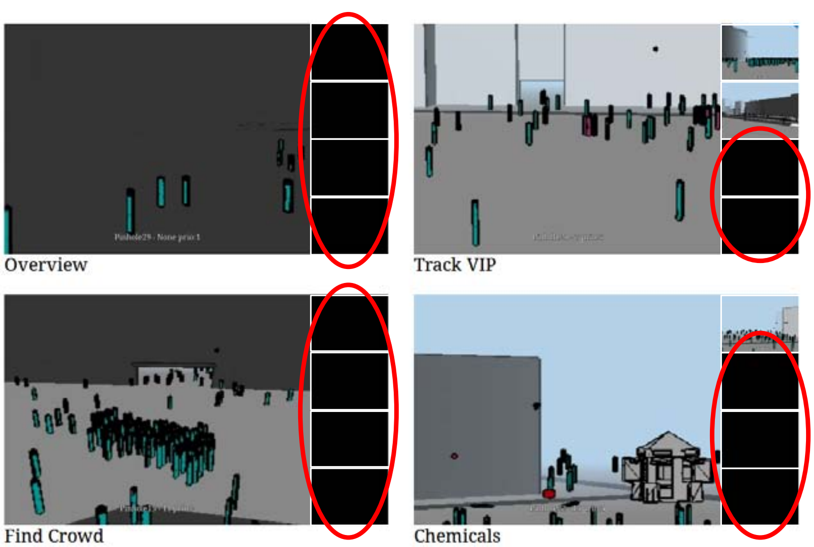
To partially address the decision-overridability concern, as shown in Figure 8, we added **live visual user feedback** in the application front-ends for C2 staff to be able to verify the system’s world model-based analysis and selection, by means of geo-map overlays showing phenomena predictions, camera streaming status and orientation, and CBRNE sensor values. The figure also shows the clickable, situation-dependent evacuation route proposals foreseen in the guidance service (example trajectories in top-right snapshot of the figure).

B. Qualitative and quantitative validation of resource use effectiveness

The platform’s prioritized camera activation results in a user experience as shown on the left side in Figure 9. Shown are the video matrices of four simultaneously active service instances (one for each planned service type planned for the scenario, as an example corresponding to Figure 7). As decided by C2 staff for the given mission, each service has been given a relative priority, e.g. VIP tracking given priority over crowd monitoring. Further, in line with mission management expert preference, the main screens (shown larger to C2 staff, and streamed at HQ stream quality) should *always* display the camera source of highest relevance to the requested service, while camera sources of less relevance to the respective services are displayed on side screens (at LQ stream quality), *as far as* possible within the actual network loading.

Without the prioritized dynamic camera activation by the service platform across all active services, random stream drops would occur at network overload, resulting not only in a bad user experience, but possibly even the drop of the video images most critical to the C2 staff decision taking.

With the dynamic priority flow reservation mechanism described in Section III however, i.e. with the service platform being able to predict based on phenomena behavioral knowledge which camera sources will become most relevant for the active services at any given time, the service platform can **rank flows according to their eventual streaming relevance**. As illustrated in the table on the right in Figure 9, the (dynamically changing and



Priority	Flow ID	QoS	Allow	Strm ID
5005	10	HQ	Yes	19
3005	10	HQ	Yes	21
2005	10	HQ	Yes	14
1005	10	HQ	Yes	28
54	9	LQ	Yes	25
53	10	LQ	Yes	6
34	10	LQ	Yes	10
33	7	LQ	No	27
32	10	LQ	No	5
24	10	LQ	No	12
23	10	LQ	No	31
14	10	LQ	No	7
13	10	LQ	No	18
12	6	LQ	No	29
11	10	LQ	No	20

Figure 9. Video user experience (left), and ranking tabel (right) for highest priority stream selection

 ISBN: 978-1-61208-421-3

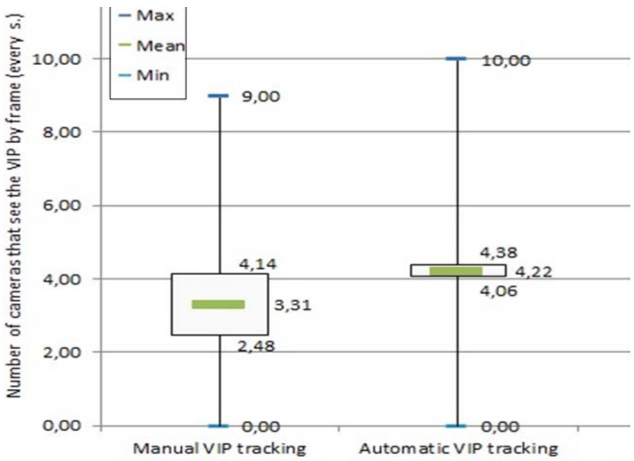


Figure 10. Automated versus human video selection

predicted) single most relevant camera source for each service (by its flow and stream identity) is assigned the highest priority (and, among each other, are ranked according to relative service priority). Camera sources with lower predicted relevance in the table, i.e. those that are estimated to be candidates for side screen display, in LQ, are dynamically priority-ranked according to their predicted relevance for each of the running services. In this way, overall, the least relevant streams are prevented from loading the network in favor of the most relevant ones. As the figure illustrates, this results in black screens (i.e. no content) to occur only for the least critical camera views. As such, **a far more acceptable user experience** is realized, fulfilling the critical mission decision support better than with a non-intelligent solution where blocking occurs randomly.

In fact, a **near-maximal hit-rate** can be seen to be reached with respect to what can be optimally obtained under the maximum overall network capacity for the given setup, as, under the mission preferences and requirements outlined in the beginning of Section IV.B., the (in general NP-hard) Knapsack optimization problem can be approximated by the simple ranking at each flow selection update cycle. (In rare instances where the phenomenon predictions would be radically wrong, there is a risk of not reserving flows properly, but the live experiments under the realistic SE-Star simulation phenomena behavior have shown that the approach appears robust against this.)

As for comparison to other stream prioritization approaches, we found Chen et al. [13] to confirm the urban security sector-relevance of concurrent tracking services, realized via a multi-service-programmable platform with a service priority notion, under the resource constraints of a WSN serving data streams from cameras, CBRNE sensors and other sensor types. Chen's approach uses a similarly elementary model for movement prediction (but considers for that an energy-friendly, distributed, cooperative tracking logic deployed across the sensor nodes) and uses a similar service-aware relevance and priority metric to dynamically determine which streams should get prioritized in the network. The main difference between both approaches lies in the fact that Chen solves resource conflicts by *controlling congestion locally* in the network nodes (favoring the most relevant streams by auction logic among the network nodes), while our approach is rather *proactively avoiding* congestion at a *global* level. Further experiments would be needed to compare the approaches in this respect, although results may be expected to be of comparable merit. While a more fine-grain optimization may be achieved in Chen's approach, at a cost, the specification flexibility of phenomenon models and service templates in our solution, and the fact that no processing is required for it on the constrained, battery-powered network nodes themselves, may however be a practical advantage in C2 or C4I mission management cases.

C. *Objective validation of operational success*

As today's C2 surveillance systems often offer only *manual* video selection and camera control, a **metric of operational success** is how well video streams are selected by the system compared to what a human operator would be able to do when presented with the same live data. Figure 10 makes this comparison for a human operation in the same VIP guidance and evacuation scenario. For obtaining the results in the figure, we conducted a validation experiment repeating the (non-deterministic) SE-Star simulation of the scenario 20-fold. Half of the runs was conducted with a human operator just activating the *person monitoring* service and observing the automated selection via the application front-end on top of the service platform (including its network stream prioritization). The other half of the runs was performed with a human operator manually selecting the virtual camera views directly in SE-Star (without network). To automate the counting of validly selected camera views during each scenario run, SE-Star was enhanced with the ability to count the number of (manually or automatically) selected cameras for which the selection was done successfully, i.e., for which the simulation indeed rendered the VIP visible to the camera view. The resulting statistics in Figure 10 show that the service platform has a higher success score in all runs, at a smaller variance. This gives a confident indication that **our automated system selection is always at least as good, and often better, than manual human selection**, for the considered type of mission scenarios. Apart from prioritizing network loading, the service platform moreover is able to *simultaneously* automate decision support for *multiple* services, tracking several different phenomena.

V. CONCLUSION

In this paper, we reported on the validation of a service platform prototype in an urban security context. We have set up experiments in such a way as to simulate the operational context for the platform as realistically as possible, using a city simulator and a constrained wireless sensor network. We found the dynamic world model-based stream prioritization across ad-hoc requested services to obtain a significant operational enhancement over current mission management tools, with respect to user experience as well as network utilization. Next to inclusion of the proposed features in product candidates, we envision generalizing on the service platform concept for broader cases and full horizontal applicability, and benchmarking it further against alternative approaches for intelligent stream selection and network resource control.

ACKNOWLEDGMENT

This work was performed in the context of the FP7 iCore project as supported by the European Commission.

REFERENCES

[1] A. Osterwalder, Y. Pigneur, G. Bernarda, A. Smith, and T. Papadakos, Value Proposition Design: How to Create Products and Services Customers Want. Wiley, 2014, ISBN: 978-1118968055.

[2] Thales Group. *Mexico City, the world's most ambitious urban security programme.* Available from: https://www.thalesgroup.com/en/worldwide/security/case-study/mexico-city-worlds-most-ambitious-urban-security-programme [retrieved: May 2015]

[3] G. K. Campbell, "Types of metrics and performance indicators appropriate to the security mission," in Measures and Metrics in Corporate Security, G. K. Campbell, Elsevier, pp. 21-56, 2014, ISBN: 978-0-12-800688-7.

[4] Open Networking Foundation. *Software-Defined Networking: The New Norm for Networks. ONF White Paper*, Apr. 2012. Available from: https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf [retrieved: May 2015].

[5] Open Geospatial Consortium. *OGC Sensor Observation Service Interface Standard Version 2.0.* Apr. 2012. Available from: http://www.opengis.net/doc/IS/SOS/2.0 [retrieved: May 2015].

[6] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. *Internet Standard RFC 3550: RTP: A Transport Protocol for Real-Time Applications*. July 2003. Available from: https://tools.ietf.org/html/rfc3550 [retrieved: May 2015].

[7] H. Schulzrinne, A. Rao, and R. Lanphier. *Proposed Standard RFC 2326: Real Time Streaming Protocol (RTSP)*. April 1998. Available from: https://tools.ietf.org/html/rfc2326 [retrieved: May 2015].

[8] J. W. Hui and D. E. Culler, "6LoWPAN: Extending IP to Low-Power, Wireless Personal Area Networks," IEEE Internet Computing, vol. 12, no. 4, pp. 37-45, July 2008, doi: 10.1109/MIC.2008.79.

[9] L. Navarro, F. Flacher, and C. Meyer, "SE-Star: a large-scale human behavior simulation for planning, decision-making and training". The 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 2015, pp. 1939-1940, ISBN 978-1-4503-3413-6.

[10] T. Winter, et al.. *Proposed Standard RFC 6550: RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks*. March 2012. Available from: https://tools.ietf.org/html/rfc6550 [retrieved: May 2015].

[11] Z. Shelby, K. Hartke, and C. Bormann. *Proposed Standard RFC* 7252: *The Constrained Application Protocol (CoAP).* June 2014. Available from: https://tools.ietf.org/html/rfc7252 [retrieved: May 2015].

[12] L. Lin and A. Liotta, "A Critical Evaluation of the IMS Presence Service," The 4th International Conference on Advances in Mobile Computing and Multimedia (MoMM 2006) Austrian Computer Society, Dec. 2006, pp. 19-28, ISBN 3-85403-215-3.

[13] M. Li, F. Ye, M. Kim, H. Chen, and H. Lei, "BlueDove: A Scalable and Elastic Publish/Subscribe Service," IEEE International Parallel & Distributed Processing Symposium (IPDPS 2011), IEEE Press, May 2011, pp. 1254-1265, doi: 10.1109/IPDPS.2011.119.

[14] L. Chen et al., "Dynamic service execution in sensor networks," The Computer Journal, vol. 53, no. 5, pp. 513-527, June 2010, doi: 10.1093/comjnl/bxp051.

Monitoring Abnormal Behavior of Hospital Patients Using RGB+D Sensors

Seung Chae and Kin Choong Yow
GIST College, Gwangju Institute of Science and Technology
Gwangju, South Korea
e-mail: chaeseung, kcyow@gist.ac.kr

***Abstract*— The ability to recognize abnormal actions or conditions of hospital patients is a very important problem as it may bring about timely medical response to a critical patient condition and may make the difference between life and death. In this paper, we propose a system that makes use of the Microsoft Kinect for Windows v2 to generate RGB+D (Red, Green, and Blue + Depth) image sequences of hospital patients. Dense 2D image features were extracted from the image sequences and then combined in a hierarchical manner to form compound features. These compound features were then mined to produce a class feature model to be used for action recognition. In the recognition phase, the RGB and the depth image data were processed separately and the responses merged to produce an overall response for action classification. Our experimental results show that this approach is able to generate good recognition rates and is comparable to other state of the art algorithms.**

Keywords-Abnormal behavior detection; learning; Kinect; data mining; Ambient Assisted Living(AAL).

I. INTRODUCTION

As many countries are rapidly becoming aging societies, significant challenges arise for the states' finances (to build more hospitals, nursing homes, etc.), as well as on the healthcare system (timely response by medical staff and post-operative care). To address these challenges, research in Ambient Assisted Living (AAL) explore modern AI (Artificial Intelligence) methods and techniques to create intelligent living environments or mobility assistants, to enable elderly or impaired people to live independently or to call for help only in an emergency.

One of the most important application environment of AAL is in the hospital. A patient is usually admitted to a hospital if he needs constant and intensive monitoring, or if he becomes too weak to move independently after a surgery. In these scenarios, a lapse in timely attention and/or response may cause the difference between life and death.

Due to the ever-increasing demand of hospital beds by the ever-increasing population of people who needs medical attention, the medical staff at a hospital always find themselves short-handed. To cope with the shortage of staff, a hospital may reduce its frequency of the medical ward rounds, leading to delays in detecting critical patient conditions. Patients are sometimes expected to call for help using the nurse call button in a non-emergency (e.g., if they need a drink of water), leading to the nurses' frustration, alarm fatigue, delays or even completely disregarding the patient's calls in real emergencies.

Due to the above reasons, AAL technologies can be very beneficial in a hospital environment. They can monitor patients in a non-intrusive manner, therefore relieving the hospital staff to perform other important duties. It can also provide non-stop, 24/7 monitoring of the patient and can even detect the case where the patient is not able to call for help nor press the nurse call button, e.g., due to temporary paralysis or laboring breath.

Recent progress in using RGB + depth sensors to detect human's action makes it an efficient and non-intrusive technology for action monitoring. In this paper, we propose a system that makes use of the Microsoft Kinect for windows v2 to detect several critical actions/conditions of hospital patients, e.g., labored breathing, failed attempts to press the nurse call button, fall off the bed, etc. When such actions/conditions are detected, an alert can be sent immediately to the medical staff on duty, greatly increasing the chance of early medical treatment and hence the survival of the patient.

The rest of the paper is organized as follows: Section II discusses some related work, and Section III describes the proposed algorithm. Section IV discusses the implementation details and Section V provides the experimental results. Section VI concludes the paper.

II. RELATED WORKS

Banerjee et al. [1] presented an approach for patient activity recognition in hospital rooms using depth data collected using a Kinect sensor. Their work detects the presence of a patient in the bed as a means to reduce false alarms from an existing fall detection algorithm. They, however, do not attempt to recognize the different actions of the patients in the hospital bed.

Saha et al. [2] studied the problem of emotion recognition from gestures using the Kinect sensor. Using the co-ordinates of joints from the upper body and the hands, a set of nine features were extracted. Using these features, they were able to uniquely identify gestures corresponding to five basic human emotional states, namely, 'Anger', 'Fear', 'Happiness', 'Sadness' and 'Relaxation'. However, in this work, the subjects were looking directly into the camera, hence it will be difficult to apply this technique to a hospital scenario.

Other techniques that address action recognition make use of depth information only or RGB information only. Ijjina et al. [3] propose an approach for facial expression recognition using deep convolution neural networks (CNN) based on features generated from depth information only. The ability of a CNN to learn local discriminative patterns from data is used to recognize facial expressions from the representation of unregistered facial images.

Gilbert et al. [4] proposed a technique that uses an overcomplete set of simple 2D corners in both space and time in the RGB data only. These are grouped spatially and temporally using a hierarchical process, with an increasing search area. At each stage of the hierarchy, the most distinctive and descriptive features are learned efficiently through data mining. This method results in fast, accurate recognition with real-time performance on high-resolution video that outperforms all other methods reported thus far in the literature.

III. Proposed Approach

A. Approach Overview

Our approach is based on the method proposed by Gilbert et al. [4] using mined hierarchical compound features. As we have additional depth data in addition to the RGB data captured by the Kinect sensor, we repeat the same process used for the RGB data on the depth data and generate two sets of association rules, one for the RGB data and one for the depth data. In the recognition process, the input RGB+D data is processed separately but is combined in the end to compute the overall support for a particular action.

Figure 1 shows an overview of our approach. In the learning process, we extract 2D Harris corners in 3 planes (*x*,*y*), (*x*,*t*), and (*y*,*t*). from the training image sequence. Each corner is then encoded and grouped within the neighborhood of a 3x3x3 spatiotemporal cube. These encoded corners are used in an iterative grouping process, which forms descriptive compound features.

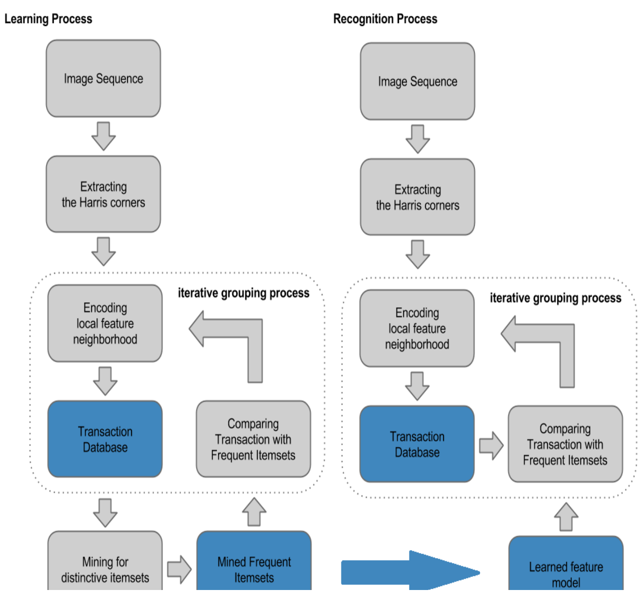


Figure 1. Overview of our approach.

Each set of grouped corners is called a Transaction, and these Transactions are collected to form a Transaction database, which is mined to find the most frequently occurring patterns, called the Frequent Item Sets. These mined item sets then become the basic feature of the next level of mining. In other words, the compound corners are re-grouped within an expanded spatiotemporal neighborhood to form a new Transaction database which data mining can be performed again. This process is repeated until we achieve a small number of complex corners at the highest level. The Frequent Item Sets of the final stage then becomes the class feature model.

The process of recognition (classification of unseen data) is almost identical with the learning part. The Harris corners of the recognition image sequence is extracted and grouped in the same way. However, instead of mining these transactions, we check to see if they match those in the learned class feature model dataset. The recognition response is obtained by summing up all the confidences for each matched transaction.

The process is run twice, once for RGB data and once for depth data. The recognition response from both is then combined to give a final recognition response value. The image sequence is then assigned an action label according to the class that maximizes the response for that sequence. In the unlikely event that no matches occur and the model score is zero, the video would be classed as not containing any action.

B. Feature Extraction

The features that we use in this work are the 2D Harris corners in the three orthogonal planes of the video spatiotemporal domain (*x*,*y*), (*x*,*t*), and (*y*,*t*). The reason that 2D corners are preferred over 3D corners (proposed by Laptev and Lindeberg [5]) is because the 3D corners may be too sparse for this approach.

The 2D corners in each of the three orthogonal planes of (*x*,*y*), (*x*,*t*), and (*y*,*t*) are detected independently using the OpenCV *goodFeaturesToTrack* function. Figure 2 shows an example of the 2D corners detected in a typical image.

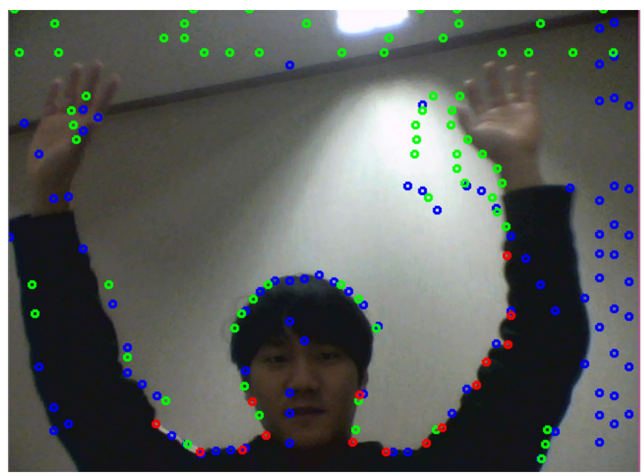


Figure 2. Harris corner detection on a frame in three orthogonal planes, (*x*,*y*), (*x*,*t*), and (*y*,*t*) [red = (*x*,*y*), green = (*x*,*t*), blue = (*y*,*t*)].

To distinguish the characteristics of each corner, we encode each one with a 3-digit number. The first digit of code represents the scale. The points of interest were detected in multiple scales, relevant to the size of the search window used. The size of the search windows we used was $\sigma_i = 3 \times 2^{i-1}$ with $i = 1, \ldots 5$, viz., 3 x 3, 6 x 6, 12 x 12, 24 x 24 and 48 x 48. This is sufficient range for the video in our experiments, where in order to achieve real-time or near real-time processing, the size of the image has to be quite small (we use 192 x108 images for our RGB data and 128x106 for our depth data).

The second digit of code represents the channel. The channel indicates where the interest pointed is detected. In this study, the channel has the value from 1 to 3 with 1 = (x,y), 2 = (x,t) and 3 = (y,t). The last digit of code represents the dominant orientation of the corner. In our system, we quantize the dominant orientation (from –180˚ to 180˚) into 8 orientations, i.e., the orientation is divided into eight equal-sized bins of 45˚ and assigned with a value from 1 to 8).

For example, a corner with code of 125 means that this corner was detected at scale 1 (i.e., 3x3 search window), from channel 2 (i.e., the (x,t) plane) and at an orientation 5 (i.e., between 0˚ to 45˚).

C. Data Mining

In our system, Association rule mining [6] is used to figure out the recurring patterns within the data. For example, an association rule has a form {A, B} → C, where A, B and C are Item sets. A, B are called the antecedents and C the consequence. This association rule example implies that if there is a customer who bought items A and B, (s)he is likely to buy item C simultaneously. To analyze each rule quantitatively, we measured a 'support' and a 'confidence' value for each rule.

To process the Transaction association rules, Agrawal and Srikant [7] developed the 'Apriori algorithm' (Figure 3). This algorithm is the best known algorithm for frequent item set mining and association rule learning over transactional databases. There are two steps in Apriori algorithm. The first step is to find all item sets that have minimum support, i.e., C_k. And the next step is to use frequent item sets to generate association 'rules' (Figure 4).

```
Algorithm Apriori(T)
    C1 ← init-pass(T);
    F1 ← {f | f ∈ C1, f.count/n ≥ minsup};
                        // n: no. of transactions in T
    for (k = 2; Fk-1 ≠ ∅; k++) do
      Ck ← candidate-gen(Fk-1);
      for each transaction t ∈ T do
          for each candidate c ∈ Ck do
              if c is contained in t then
                      c.count++;
          end
      end
          Fk ← {c ∈ Ck | c.count/n ≥ minsup}
    end
    return F ← ∪k Fk;
```

Figure 3. The Apriori algorithm

```
Algorithm RuleGeneration(X)
 For each frequent itemset X,
 For each proper nonempty subset A of X,
    Let B = X - A
    A → B is an association rule if
        Confidence(A → B) ≥ minconf,
        support(A → B) = support(A∪B) =
            support(X)
        confidence(A → B) = support(A ∪ B) / sup-
port(A)
```

Figure 4. The Rule Generation algorithm

To clarify the concept of Association rules and Apriori algorithm, let's take an example from a real market basket analysis. Let $I = \{i_1, i_2 ,..., i_m\}$ be a set of items (an 'item' can be a real item in a basket and I is the set of all items sold in the store). Transaction t is a set of items, where $t \subseteq I$ (a transaction can represent items purchased in a basket), and the Transaction database T means a set of transactions, which can be written as $T = \{t_1, t_2 ,..., t_n\}$. An association rule is an implication of the form: X → Y, where X, Y are item sets.

D. Learning

To learn the frequent mined corner configurations (i.e., the class feature model), we apply the same method of neighborhood encoding as Gilbert et al. [4] for all the features that were detected in the feature extraction stage. In the encoding scheme, a regular 3 x 3 x 3 grid is used to establish a neighborhood for encoding the relative position of corners.

Figure 5 shows four corners that have been detected in the region around a central corner that is marked with a red cross. For the neighborhood encoding, we are interested in the relative positions of each corner relative to the center corner. In a 3 x 3 x 3 grid, there are a total of 27 cells. Each cell is numbered from 0 to 26, starting from the smallest value of t, y and x an increasing first in x, then in y and finally t. For example, in Figure. 5, the top-left corner of the front 3 x 3 grid (i.e., at t-ω) will have a position code of 00 and the bottom-right hand corner of the back 3 x 3 grid (i.e., at t+ω) will be 26. The center cell will therefore be 13.

Each corner has its individual three-digit code, and in the neighborhood encoding, each corner is then prefixed with an integer that denotes the grid cell where it occurs. For the central corner in Figure 5, the cell number is 13, and hence, the center feature is represented by the five-digit number 13125.

This five-digit number is known in data mining as an item and encoding all of four corners in the grid will yield four items, e.g., 00321, 08237, 13125 and 20112. The items are then concatenated into a large 1D vector, known within the mining community as a Transaction vector. Here T = {00321, 08237, 13125, 20112}. For the purposes of the training stage, each Transaction vector is appended with the label of the associated action class, α. Hence, the Transaction vector becomes {00321, 08237, 13125, 20112, α}. This encoding process is then repeated for all 2D corners detected in the video sequence to produce D_1, the transaction database for the first stage of mining.

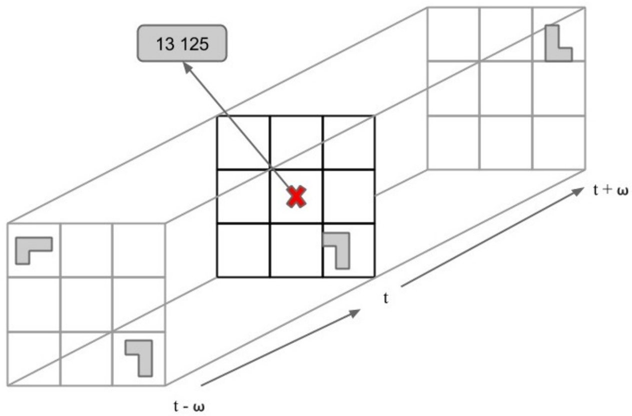


Figure 5. The grid centered on a corner shown by a cross. Four other corners are found within the neighborhood defined by the grid.

The process is then repeated at a second level l = 2, where the compound feature detected in level l = 1 becomes the input to the next higher level of hierarchical grouping. Using the same encoding scheme of level l = 1, we prefix each of these compound feature with the cell position that it was found. At these higher levels, we still use a 3 x 3 x 3 grid with 27 cells, but each cell now has a larger size, where ω^l; = 2 x ω^{l-1}. In our experiments, we use l from 1 to 3.

E. Recognition

After the training has taken place, the frequently recurring distinctive and descriptive compound features for each class α, M(α), are produced,. To classify an unseen RGB video sequence, we use the same procedure as Gilbert et al. [4]. The video is analyzed in the same way as in the learning process, but instead of mining patterns from D, only patterns that exist in M(α) are passed to the next level. The confidence of each transaction in M(α) is used to weight the matches, as a high confidence would indicate that the Transaction T is distinctive compared to other classes. The use of the confidence ensures that if the transaction is matched with several classes, the confidence will provide a measure of the discrimination between those classes. The response R of the classifier is given by

$$R_\alpha = \frac{1}{|D \cap M(\alpha)||M(\alpha)|} \sum^{\forall T_i \in D} m(T_i, M(\alpha)), \quad (1)$$

where

$$m(T_i, M(\alpha)) = \begin{cases} conf(T_i \Rightarrow \alpha), & T_i \in M(\alpha) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The entire process is then repeated for the depth image sequence. The two responses from the RGB and the depth data are then combined to give an overall response. The image sequence is then assigned an action label according to the class that maximizes the response for that sequence. In the unlikely event that no matches occur and the model score is zero, the video would be classified as not containing any action.

IV. IMPLEMENTATION

A. Hardware/Software specifications

In our implementation, we make use of the Microsoft Kinect for Windows v2 that was released in the summer of 2014. The Kinect v2 comes with a 1080p color camera and a 512x424 depth sensing camera [8]. The Kinect was connected to an Intel Core i5-3210M 2.50GHz laptop running Windows 8.1, Microsoft Visual Studio 2013 and OpenCV 2.4.10.

B. The six action classes

In our experiments, we have identified a set of six actions that we want to recognize. Two of the action class (*Still.* and *Rolling*.) represent normal action (or normal condition) of a patient on a hospital bed.

1) *Still.* This is the default normal action. The patient i s either sleeping or lying motionless on the bed.

2) *Rolling*. This is also a normal action. The patient turns in the bed occasionally.

3) *Coughing*. This is the first abnormal action. The patient is coughing violently for a prolonged period of time, indicating a serious deterioration of the patient’s condition.

4) *Bottle.* This represents a patient’s repeated but failed attempts to reach for a water bottle on the bedside table. The patient requires water but is too weak to reach for it by himself.

5) *Button*. This represents a patient’s repeated but failed attempt to reach for the nurse call button on the wall beside the bed. The patient could be in a life-threatening situation but is too weak to reach the button.

6) Falling. This represents the patient completely falling off the bed, which may be caused by excessive turning in bed, or when the patient reach out too far for an object (e.g., water bottle).

The default action (i.e., Still) is necessary because our algorithm will always assign an action to the test image sequence (the action class that have the largest response). If lying still is not considered as an action, then the algorithm will need some nontrivial modification to handle them so that it will not cause the unintended rising of the false positive rate.

Figure 6 shows examples of the RGB and depth image of each of the six action class.

C. Generating the Image Database

For each action except the first (i.e., lying still), we capture about 5-8 seconds of the action (at 30 frames per second) in both RGB and Depth images at full resolution (RGB: 1920x1080, Depth: 512x414). Each action was preceded by about 1.5 to 2 seconds of no action (i.e., lying still) which actually constitutes Action 1 (Still).

Due to the computation complexity of the algorithm, we have rescaled the RGB images by a factor of 10 (to 192x108 pixels) and the Depth images by a factor of 4 (to 129x106 pixels) in our experiments.

 ISBN: 978-1-61208-421-3

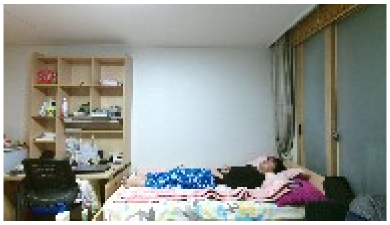
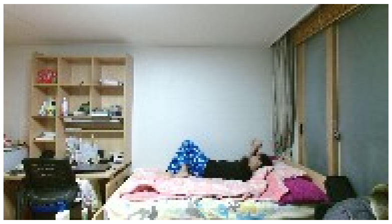
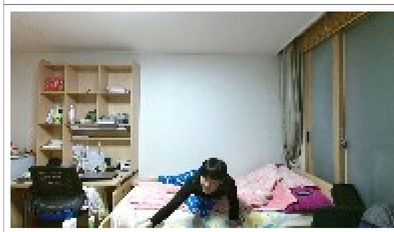
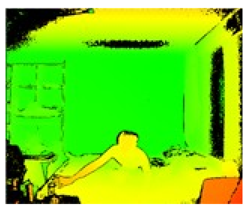
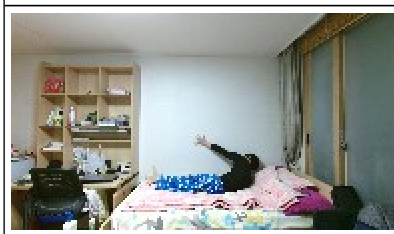
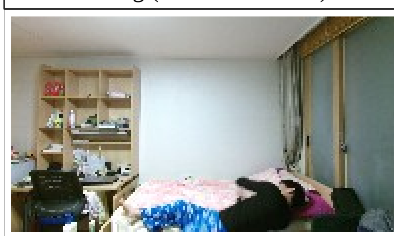
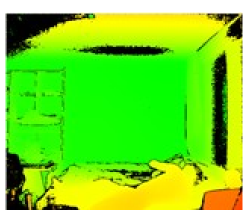
Action name	No. of frames
Still	220 frames
	
Rolling	168 frames
	
Coughing (Abnormal action)	210 frames
	
Bottle (Abnormal action)	219 frames
	
Button (Abnormal action)	215 frames
	
Falling (Abnormal action)	204 frames
	

Figure 6. Example RGB+D images of the six action classes.

V. Experimental Results

A. *Recognition Rate*

We test our system with 52 test sequences (of 2 second duration each, i.e., 60 frames). The results are shown in Table I.

TABLE I. Recognition Rates

Action	No. of Test Sequences	No. of Successful Recognitions	Recognition rate(%)
Still	4	4	100.0
Roll	6	5	83.3
Cough	11	9	81.8
Bottle	8	6	75.0
Button	13	11	84.6
Fall	10	10	100.0

B. *Confusion Matrix*

Figure 7 shows the confusion matrix of our experiments.

	Still	Roll	Cough	Bottle	Button	Fall
Still						
Roll	17					
Cough				9	9	
Bottle			13		12	
Button			8	7		
Fall						

Figure 7. Confusion matrix

C. *Comparison with other methods*

We compare our method with traditional classification algorithm such as binary decision tree, ensemble tree, k-NN (k-Nearest Neighbors algorithm), SVM (Support Vector Machine) with radial basis function kernel and neural network classifier with back-propagation learning.

TABLE II. Comparison with other methods

Method	Average classification accuracy (%)
Binary decision tree	76.63
Ensemble tree	90.83
k-NN	86.77
SVM with radial basis function kernel	87.74
Neural network classifier with back-propagation learning	89.26
Our method	86.54

The results are shown in Table II. We observe that our system is comparable to other state-of-the-art methods.

VI. Conclusion and future work

In this paper, we have proposed a system that makes use of the RGB+D sensors available on the Microsoft Kinect for Windows v2 to determine abnormal behaviors of hospital patients. Our proposed system processes the RGB information and Depth information separately, and then combines the responses to make the action classification.

In the learning phase, for each of the RGB and Depth image sequence, 2D Harris corners are detected in each of the three orthogonal planes of the video spatiotemporal domain (x,y), (x,t), and (y,t) and then encoded based on their relationship in an expanded spatiotemporal neighborhood. These encoded corners formed Transactions in a Transaction Database which is subsequently mined to obtain the class feature models. Two separate transaction databases were kept for the RGB data and the depth data.

During the recognition phase, a similar process of extracting 2D Harris corners from the RGB test sequence and the Depth test sequence is used. These corners were also encoded into Transactions but instead of mining them, they are matched with the class feature model. The recognition response is obtained by summing up all the confidences for each matched transaction.

Our experiments showed that this approach is able to generate good recognitions rates and is comparable to other state-of the art algorithms. Our future work will include collaborating with an actual hospital to perform live trials on their patients.

References

[1] T. Banerjee, M. Enayati, J. M. Keller, M. Skubic, M. Popescu, and M. Rantz, “Monitoring patients in hospital beds using unobtrusive depth sensors” in Conf Proc IEEE Eng Med Biol Soc. 2014, 2014, pp. 5904-5907.

[2] S. Saha, S. Datta, A. Konar, and R. Janarthanan, “A Study on Emotion Recognition from Body Gestures Using Kinect Sensor”, International Conference on Communication and Signal Processing, 2014, pp. 56-60.

[3] E. P. Ijjina and C. K. Mohan, “Facial Expression Recognition Using Kinect Depth Sensor and Convolutional Neural Networks”, 13th International Conference on Machine Learning and Applications (ICMLA), 2014, pp. 392 – 396.

[4] A. Gilbert, J. Illingworth, and R. Bowden, , “Action Recognition Using Mined Hierarchical Compound Features”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 33, Issue 5, 2010, pp. 883- 897.

[5] I. Laptev and T. Lindeberg, “Space-Time Interest Points,” Proc.IEEE Int’l Conf. Computer Vision, 2003, pp. 432-439.

[6] R. Agrawal, T. Imielinski, and A. Swami, “Mining Association Rules between Sets of Items in Large Databases,” Proc. 1993 ACM SIGMOD, 1993, pp. 207-216.

[7] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules in Large Databases,”, Proc. 20th Int’l Conf. Very Large Data Bases, 1994, pp. 487-499.

[8] Microsoft Kinect for Windows v2 features [Online]. Available at http://www.microsoft.com/en-us/kinectforwindows/meetkinect/features.aspx. Retrieved 23 Mar 2015.

Flexible Management of Ambient Assisted Living Environments by Ontology-Based Configuration

Tom Zentek, Jürgen Bock, Asarnusch Rashid

FZI Forschungszentrum Informatik
Karlsruhe, Germany
Email: [zentek|bock|rashid]@fzi.de

***Abstract*—Ambient Assisted Living (AAL) environments are Smart Environments supporting elderly people and their caregivers in their daily life. Such environments are characterised by a wide heterogeneity of software, hardware, and human components as well as rapidly changing needs of the assisted persons. Therefore, flexible management is a crucial success criterion for AAL environments that has not been addressed in the state of the art of AAL research yet. In this paper, we propose an ontology-based configuration framework that fills this gap. Using ontologies to describe AAL environments enables a common understanding and a reusable semantic representation of AAL use cases. This improves the flexibility needed to configure the environment in order to deal with changing individual needs of the assisted persons.**

***Keywords*–ambient assisted living; configuration; semantic; knowledge-based configuration; OWL; smart environment; ambient intelligence**

I. INTRODUCTION

With the increasing development of Internet of Things technology, the vision of Smart Environments becomes real. Smart Environments consist of hardware (e.g. sensors) and software services linked together and embedded "in everyday settings and commonplace tasks" [1]. Ambient Assisted Living (AAL) is an application of Smart Environments, dealing with the assistance of elderly peoples and their caregivers in their daily life. AAL hardware is installed in the environment of the assisted persons. An *AAL platform* collects the data from the connected hardware and provides software services to process the collected data. Also, it controls the hardware and interacts with the assisted persons. The different *components* from the AAL environment e.g. software, hardware and human resources can be grouped into *AAL use cases*. End users of these systems are *assisted persons* and their caregivers. *Deployers* provide end user support in installation and maintenance of the AAL environment. *Developers* are responsible for the AAL software and AAL hardware.

Regarding the management of an AAL environment, two factors are mainly challenging:

Heterogeneity of AAL solutions: In everyday settings end users will use AAL components from different developers in their home environment, but they want to use the same hardware resources and share data among different software services. So far, manufacturers and technology providers develop their own standalone solutions, each of them having their own unique data representation.

Individual and changing needs of the users: Users have individual settings (e.g. home environment, familiar circumstances) and individual demand for assistance (e.g. mental state, physical condition). Furthermore, the elderly have rapidly changing needs, depending on their current state of health and general condition.

These two challenges are addressed by the flexibility of an AAL environment as a "measure of how easily a system's capabilities can be modified in response to external change" [2]. Therefore, we conclude that flexibility is a crucial success criterion for the management of AAL environments.

Previous work does not address this gap yet: Due to the consolidation and enhancement of a variety of AAL frameworks that had been developed in research projects over the last years [3], [4], [5], [6], one can suggest that today semantic middlewares have established themselves. They use formal languages such as Resource Description Framework (RDF) , Web Ontology Language (OWL) and Business Process Execution Language (BPEL) to gather declarative knowledge, service requests as well as procedural knowledge in a machine readable form and to realise use cases. So far, this formal description has been a very suitable but static basis for configuration aspects. In literature, one can find approaches for ontology-based configurations in other domains. Ardito et al. [7] describe how orthogonal data sources can be connected for configuration purposes by means of ontologies and mapping processes. Dong et al. [8] attempt to simplify configurations with the help of reasoning based on a semantic description. However, none of these approaches in their present form are directly applicable to management of an AAL environment and the above described challenges. Further preliminary work can be found in the field of smart homes. For instance, there are publications [9], [10] about the configuration and personalization of intelligent houses adjusted to the residents requirements. However, semantic approaches and a changing medical context in terms of AAL have not been addressed so far.

Ontology-based configuration as a specific implementation of knowledge-based configuration [11] provides a possible solution to fill this gap. The two core advantages of an ontology [12] used in this approach are: (i) Reusable ontology concepts can ensure the interoperability of the different use case components. (ii) They provide a common understanding for the knowledge representation in the environment and its stakeholders.

Therefore, this paper investigates the application of an

 ISBN: 978-1-61208-421-3

ontology-based configuration to increase flexibility of the management of AAL environments and contributes to the state of the art by presenting an ontology-based configuration framework. This approach allows for flexibility in the phases [13] of development, design, installation, customisation and maintenance of the implemented use cases including their components: software, hardware and human resources. Furthermore, some of these phases can be automated and all of them can be tool-supported.

As an example, think of an elderly person getting up from bed at night to have a glass of water. The use case "fall-prevention" uses a controlling software component processing data from motion detection sensors and actuators to light the way to the kitchen. In our approach the light source would be a general concept reusable for managing all lights on the way to the kitchen or even in the house. The flexibility gained by our approach can be shown in this example by considering that different motion sensors and light switches from different hardware manufacturers are installed in the environment. With increasing dementia of the resident it is necessary to integrate a second use case. It allows the monitoring of the usage of the front door to ensure his or her safety. Therefore, the information about his current location in the environment is needed and a sensor to detect the door state. The already installed motion detection sensors can now be reused for the new use case and a contact sensor has to be mounted at the front door. Now two services ("fall-prevention" and "front door control") are interoperable using the same hardware based on the same reusable ontology concepts.

This paper is structured as follows. Section II describes the requirements for the AAL environment management, collected by a stakeholder analysis. Based on this input, Section III demonstrates the developed configuration framework with the according middleware, ontologies, processes and tools. Subsequently, Section IV presents the implementation of the framework. A discussion of the results concludes this paper in Section V.

II. Configuration Requirements in AAL Environments

The requirement analysis is divided into a stakeholder analysis and subsequently into expert workshops. For the stakeholder analysis, results of the standardisation group AALIANCE [14] are used to identify the main stakeholders in the management of an AAL environment. Second, a workshop with domain experts (representing the identified stakeholders, 10 participants) was organized to gather their configuration requirements. In a structured discussion guided by the deployment process from five other AAL platforms (Persona, Soprano, MPOWER, OASIS, AMIGO) the experts finally came up with a set of stakeholders and requirements.

Five different groups of stakeholders [14] are involved directly into the phases (development, design, installation, customisation, maintenance) of the configuration of an AAL environment. The following list gives an overview about the defined subset of stakeholders which are directly involved in the configuration. The group of developers is split in two main roles: 1. The *software developer* implements the application; 2. The *hardware developer* manufactures the devices used by the software. After purchasing a software, hardware or acquiring human resources components, the group of deployers ensures the integration of the components in the AAL environment. The majority of this group (assisted persons) is not directly faced with the configuration process, but has to be involved as they need to provide information about themselves and want the AAL environments to be configured for their needs.

Altogether the expert group came up with 63 technical requirements [15]. They can be summarised to the following four main requirements:

- R1. Easy Integration of new functionality: A flexible management shall allow for easy integration of new software, hardware, and human resources.
- R2. Shared Collection and Provision of Information: A flexible management shall enable the stakeholders to share information about the AAL environment between each other in a common way.
- R3. Adjustment to End User: A flexible management shall allow for a comfortable way to adapt different software, hardware and human resources to a wide variety of different end user and their needs.
- R4. Maintenance of the AAL System: It should be possible to maintain the installed software, hardware and human components easily.

III. Configuration Framework

To address the above mentioned four requirements R1, R2, R3 and R4 a configuration framework was developed based on a semantic middleware universAAL [16]. It is implementing the AAL platform of the AAL environment. With such a middleware it is possible to use ontologies within different use case components. The ontologies allow for reusing knowledge about the AAL environment implemented by another use case component. Together with the common understanding between components, this is the base to make heterogeneous components interoperable.

For the organisation of the AAL environment management a configuration process is developed. It defines the phases influencing the configuration of an AAL use case beginning with its development and ending with its maintenance. Tools are developed allowing for editing special parts of the ontologies and to assist the stakeholders through the configuration to provide an easy and powerful AAL environment management. The development of the ontologies, the configuration process and the tools followed an iterative approach. In the first iteration, six participants in a focus group defined the first draft of the ontologies and the configuration process. For a proof of concept at the end of this iteration, a first implementation of the configuration framework (named "Jambi") was developed and installed in five living labs)CIAMI Living Lab TSB (Spain), CERTH-HIT und CERTH-IBBR Labs (Greece), Laboratory of Environment Intelligence UPM (Spain), Smart Home AIT (Austria), FZI Living Lab AAL (Germany)) [17], [18]. These experiences were used in the second iteration to revise the ontologies and the processes with seven participants in a focus group and to implement a second prototype (named "Vaadin"). Additionally, the second step involves creating tool support for the stakeholders to ease the usage of the configuration framework. Finally, the second prototype is currently tested in nine real homes (Valencia (n=5), Madrid (n=3), Vienna (n=1)).

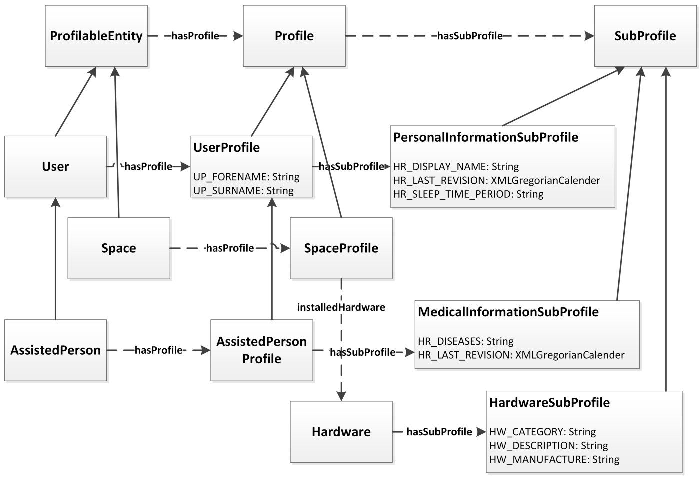


Figure 1. Configuration related ontologies.

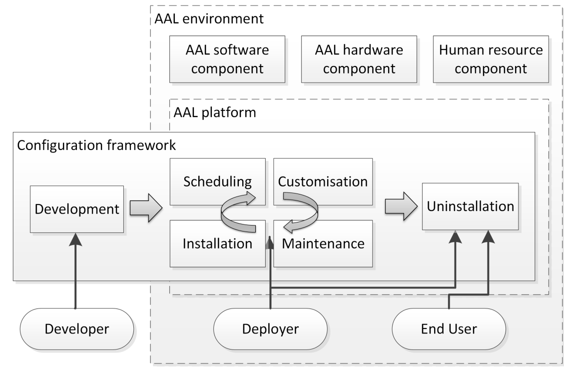


Figure 2. The configuration framework with its stakeholders and their involvement in the configuration phases.

A. *Profile Ontology*

To ensure the interoperability between use cases the information about its components is represented in ontologies. They are delivered with the use case component and integrated in the AAL environment with the configuration framework. In order to extend the middleware ontology concepts with a new ontology, coming for software, hardware or human resource components, one has to follow design patterns. On the highest layer the middleware ontology provides three concepts (cf. Chapter 1): Every use case component (software, hardware or human resource) extends the `ProfilableEntity` in order to create new sub-concepts for its own usage. Every `ProfilableEntity` (Domain) can have one `Profile` (Range). Their instances are connected via `hasProfile` (Property) to represent general Information about the software, hardware or human resource components. Every `Profile` (Domain) can have several `SubProfiles` (Range). Their instances are connected by `hasSubProfile` (Property) to represent special information needed by the components for the use case purposes.

For the example use case “fall-prevention” the ontology has to be extended with concepts in the following way (cf. Figure 1). For the human resource part a new concept of an assisted person is needed which automatically inherits the `PersonalInformationSubProfile`. This sub-profile implements a data property to define the sleeping time period (time in which the lights have to be switched on at night). All used functionality of the motion detection sensors and light switch actuators is modelled in the `HardwareSubProfile`. For new concepts, for example a picture of the hardware, a new data property linking to the picture or even a whole new `SubProfile` would be created (e.g. related to the manufacture (`HardwareManufactureSubProfile`) or the communication protocol). The concept of spaces is used here to group hardware related to different locations.

Furthermore, for the second use case with the increasing dementia a new sub profile (e.g. `MedicalInfomationSubProfile`) can be integrated and reuse previously defined ontology parts.

B. *Process and Tools*

To have the use case working on the AAL platform, configuration is needed, before processing a light switch request. The configuration framework splits into five phases for the integration of the use case in the AAL environment. Figure 2 gives an overview of these phases and their alignment to the stakeholders directly involved in the configuration (cf. Chapter II). They are overlapping in the phases if some interaction could be necessary during the use case integration. The configuration framework is implemented on top of the AAL platform (see Figure 2). The framework deploys the software, hardware and human resource components on the AAL platform though these phases:

The *development* phase contains all steps building software, hardware and human resource components.The two main stakeholders are the software and hardware developers. They implement the components and make them configurable. Therefore, the developers have to deal and extend with three types of ontologies (software, hardware, human profile; cf. Chapter III-A).

The use case “fall-prevention” can easily extend the ontology concepts to its needs (cf. requirement R1, R2, R3). In the configuration phase a tool is provided that allows for an easy packaging of the develop components to one use case. The tool also adds deployment information in a manifest.

The *design* phase is the place to find and combine software, hardware and human resources components to the needs and preferences of the elderly. The case manager uses the use case descriptions from the developers and the ontologies to combine different components to meet the requirements. This ensures a high flexibility and fault tolerance of the AAL platform.

The case manager can now easily design the AAL environment using the motion detection sensors (from the “fall-prevention” use case) also for the second “front door monitoring” use case (cf. R1).

The *installation* phase integrates the different components of the use case in the AAL platform. During the installation three different stakeholders are involved.

The case manager mainly coordinates the installation. For this task he is supported by the administrator of the AAL environment. During the installation, the use case manifest links to several information items (e.g. licences) and guides through the installation process. Based on the manifest information, the compatibility of the use case to the rest of the system is checked. Also, the manifest describes the requierd information to install the ontologies delivered with the use case in the platform (e.g. path of the ontology profile, ID of the runtime instance loading the ontology). After the new ontologies are matched to the main system ontology and all dependencies are solved, the new ontology are resolved successfully.

The technician is responsible for the installation of hardware components. Mostly he will integrate new sensors and actuators in the environment and ensures that they are known in the AAL platform based in their semantic description. If the ontology concepts coming from the use case components matching the requirements, they can be added in the platform automatically. This phase addresses requirement R1 and R2.

The *customisation* adjusts the software, hardware and human resources components to the elderlys preferences. This is mostly done by creating instances of the previously installed ontologies. The installed hardware needs a representation on the software level to be used in the system. Also, other software components and human resources are set-up. Some of the information in the AAL environment can be reused in this step or can be used to automate the configuration by its machine readable description. Based on this it is possible to create fall back configurations to be fault tolerant.

For the example use cases the motion detection sensors and the light switch actuators have to be instantiated in the ontology. The case manager, possibly with help of the technician will ensure to set them up according to the ontology concepts. They will set the location, a label and other needed parameters. Furthermore, they may have to fill in additional information about the assisted person. After the hardware and human ontology is instantiated it could be necessary to link the software part to right hardware concepts or instances. All these information will adjust the use case to the assisted person (cf. R2, R3). Especially for the second use case reusing the motion sensors this is necessary (cf. R1).

A tool to creating or editing the ontology instances and to initialise the software part is located in the configuration framework.

The *maintenance* enables the deployer to easily control the environment. In this phase of the configuration process it is possible to edit software, hardware and human resources components by changing ontology instances. Of course, it is also possible to delete or uninstall components or even whole use cases. The maintenance phase can detect errors in the AAL platform and react accordingly to the fall back configurations or/ and inform the maintenance provider.

In the "fall-prevention" use case a ceiling light could be broken. With the common understanding of light sources and being organised in hierarchical concepts the platform and the maintenance provider can reconfigure the use case to use another light source located in the kitchen (cf. R4).

IV. IMPLEMENTATION

The configuration framework was implemented in Java, OSGi (Open Services Gateway initiative: http://www.osgi.org/) and a GWT (Google Web Toolkit: http://www.gwtproject.org/) derivate called Vaadin. It is part of the universAAL middleware running in the AAL environment accessible via browser from the outside. The first phase of the configuration framework is not part of the implementation because it provides guidance about developing configurable use cases and ontologies. The developer will just use his preferred IDE. The universAAL project provides Eclipse plug-ins to transform ontologies to Java and vice versa, to generate configuration files and to package the use case with its components and manifest. The implementation of the framework supports the installation, customization and maintenance phases. The implementation of the configuration framework consists of three main managers: The *Deploy Manager* installs and uninstalls use cases on the AAL platform is the purpose of the Deploy Manager. The *Configuration Manager* configures the AAL-service based on the configuration file delivered with the use case. The *Ontology Manager* enables the case manager to create or edit ontology instances (cf. Chapter III-A).

V. DISCUSSION

Based on a detailed stakeholder analysis, a configuration framework was developed in order to flexibly manage AAL environments. It is integrated on the semantic middleware of the universAAL platform. This framework contains (i) a process guiding the stakeholders through the use case configuration and use case ontology development and (ii) tool support for the different phases in the process. With the configuration framework, the flexible management of AAL environments is supported.

This framework was installed and used in five living labs in Europe and in nine real homes. According to the iterative approach of the framework development, the first framework prototype "Jambi" was installed and tested in five living labs. The following example of the use case "nutritional advisor" with its ontology extensions shows how the evaluation was performed: Each living lab was asked to deploy this use case with the configuration framework:

1) Search and purchase the "nutritional advisor" use case in the online store
2) Download and install the use case
3) Customise the use case according to the needs of the assisted person
4) Verify correct integration of the use case in the AAL environment

All five living labs succeeded in deploying the use case with an average time of 36 minutes (4,80 standard deviation). One of the case managers deploying the use case needed support from the developer of the "nutritional advisor". The big time difference is caused by the varying prior knowledge of the case managers. Some of them had already deployed AAL use cases. They all agreed that the deployment process was too complex: "install an app involves a lot of steps (too many clicks!!!)" but they also agreed that it was helpful to have the guidance by the configuration framework. The goal in the next iteration was to try to automate configuration steps. Based on the problem that developers do not know which ontologies are available in the AAL environment, they started to deliver more and more own ontology extensions with their use case components. With the ontology editors help the deployer can easily reuse ontologies and their instances during the configuration. Of course, developers outside of the AAL environment were not aware of ontologies already deployed in the environment. After the first evaluation the refined framework was implemented in the second prototype “Vaadin”. It was installed with the universAAL middleware in nine real homes running six use cases with 48 ontology extensions. The field trial can be seen as a proof-of-concept for the configuration framework. At the moment it is still on-going.

In the near future, the framework will be used in in the European research project ReAAL [19] which has the goal to deploy the universAAL platform with the configuration framework in 7000 real homes in Europe. The implementation of the framework the ontologies are available open source [20].

References

[1] D. Cook and S. Das, Smart Environments: Technology, Protocols and Applications. Wiley, 2004.

[2] E. T. Ryan, D. R. Jacques, and J. M. Colombi, “An ontological framework for clarifying flexibility-related terminology via literature survey,” Systems Engineering, vol. 16, no. 1, Mar. 2013, pp. 99–110. [Online]. Available: http://doi.wiley.com/10.1002/sys.21222

[3] P. Wolf, “openAAL - the open source middleware for ambient-assisted living (AAL),” in Proceedings of the AALIANCE conference, 2010.

[4] M. Mikalsen, “Interoperability Services in the MPOWER Ambient Assisted Living Platform.” in Vol. 150 IOS Press, 2009, pp. 366–370.

[5] A. Schmidt, “SOPRANO Ambient Middleware: Eine offene, flexible und marktorientierte semantische Diensteplattform fr Ambient Assisted Living.” in Proceedings of the 2. Deutscher Kongress Ambient Assisted Living, 2009.

[6] A. Fides-Valero, “The PERSONA Framework for Supporting Context-Awareness in Open Distributed Systems.” in Proceedings of European Conference on Ambient Intelligence, 2008.

[7] C. Ardito, B. Barricelli, P. Buono, M. Costabile, R. Lanzilotti, A. Piccinno, and S. Valtolina, “An Ontology-Based Approach to Product Customization,” in End-User Development. Springer, 2011, vol. 6654, pp. 92–106. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21530-8_9

[8] M. Dong, D. Yang, and L. Su, “Ontology-based service product configuration system modeling and development,” Expert Systems with Applications, vol. 38, 2011, pp. 11 770–11 786.

[9] C. Cetina, P. Trinidad, V. Pelechano, and A. Ruiz-Cortes, “Customisation along Lifecycle of Autonomic Homes,” in 3rd International Workshop on Dynamic Software Product Line (DSPL09), 2009.

[10] F. Kawsar, T. Nakajima, and K. Fujinami, “Deploy spontaneously: supporting end-users in building and enhancing a smart home,” Proceedings of the 10th international conference on Ubiquitous computing (2008), Sep. 2008, pp. 282–291. [Online]. Available: http://dl.acm.org/citation.cfm?id=1409635.1409673

[11] M. Stumptner, “An overview of knowledge-based configuration,” AI Communications, vol. 10, no. 2, Apr. 1997, pp. 111–125. [Online]. Available: http://dl.acm.org/citation.cfm?id=1216064.1216069

[12] T. R. Gruber, “Toward principles for the design of ontologies used for knowledge sharing?” International Journal of Human-Computer Studies, vol. 43, no. 5-6, Nov. 1995, pp. 907–928. [Online]. Available: http://dl.acm.org/citation.cfm?id=219666.219701

[13] T. Zentek, A. Marinc, and A. Rashid, “Methods and Tools for Ontology-based Configuration Processes of AAL Environments,” in Ambient Assisted Living, R. Wichert and H. Klausing, Eds. Berlin Heidelberg: Springer, 2013, pp. 213–224.

[14] G. van den Broek, F. Cavallo, and C. Wehrmann, “AALIANCE Ambient Assisted Living Roadmap,” Tech. Rep., 2010. [Online]. Available: http://books.google.com/books?id=dDXPKWN61msC&pgis=1

[15] T. Zentek, C. Yumusak, C. Reichelt, and A. Rashid, “Which AAL Middleware matches my requirements? An Analysis of Current Middleware Systems and a Framework for Decision-Support.” in Ambient Assisted Living AAL, W. Reiner, Ed. Berlin Heidelberg: Springer, 2014.

[16] W. P. Tazari M.-R,Furfari F,Fides Valero Á,Hanke S,Höftberger O,Kehagias D,Mosmondor M,Wichert R, “The universal Reference Model for AAL,” in Handbook of Ambient Assisted Living, J. et al. Augusto, Ed. IOS Press, 2012, pp. 610 – 625.

[17] B. Bergvall-kåreborn, B. Bergvall, M. Holst, and A. Stå hlbröst, “Concept Design with a Living Lab Approach,” Proceedings of the 42nd Annual Hawaii International Conference on System Sciences, HICSS (2009), 2009, pp. 1–10.

[18] V.-p. Niitamo, S. Kulkki, M. Eriksson, and K. A. Hribernik, “State-of-the-Art and Good Practice in the Field of Living Labs,” 2006.

[19] Project ReAAL, “ReAAL,” http://www.cip-reaal.eu/home/, 2015, [Online; accessed 19-Juni-2015].

[20] Project uAAL, “universAAL,” http://forge.universaal.org/svn/uaaltools/trunk/uCC/ ontology.universaal.org, 2015, [Online; accessed 19-Juni-2015].

Geolocation Architecture for Combining Multiple Technologies

Jean-Baptiste Drucker   Tony Ducrocq   Colombe Herault   Guillaume Lefebvre   Eric Ptak

Worldline, R&D User eXperience
Lille, France
Email: name.surname@worldline.com

***Abstract*—Industry, retail and logistic markets are more and more asking for geolocation solutions, especially indoor ones. They often have different needs at the same time and require several different technologies. Until now, there was no solution for combining multiple technologies. To meet these needs, we propose a distributed, multi-user, multitechnology architecture for combining different geolocation solutions.**

Keywords–Indoor-location; middleware; distributed;

I. INTRODUCTION

Context-aware computing, in particular based on geolocation, is a big trend in the field of mobile computing. The market of location-based services is mainly driven by retailers, who want to improve their Customer Relationship Management by providing services such as couponing in the best place at the best time. But it is also boosted by the health market and Business to Business applications, such as those for maintenance and traceability uses.

Recently, cheap, easy-to-deploy technologies such as iBeacon [1] have been promoted, but lots of other technologies, such as Wi-Fi fingerprinting [2], can provide valuable solutions for geolocation. Nevertheless, the development of geolocation-based applications depends very much on the chosen technology. Therefore, it is not possible to use a unique application in several places if they are not equipped with the same location system. For example, if a retailer manages several shops, one that is equipped with iBeacon, another with inaudible sound, and the last one with Wi-Fi fingerprinting, then three different implementations of the localization application have to be developed.

Moreover, it is not easy to combine technologies to use them concurrently and provide enhanced services that dynamically take advantage of all the equipment present. For instance, in hospitals, parts of buildings, such as Computer Tomography scanner rooms, are equipped with accurate, highly available location systems that are based on active Radio Frequency Identification (RFID) [3] or Wireless-Fidelity (Wi-Fi) tags [4]. They aim at locating patients, healthcare equipment and patient records in order to improve medical furniture usage and reduce the number of medical errors. But these hardware systems (antennas and tags) are quite expensive and hard to deploy. Therefore, only critical rooms are equipped with them.

On the other hand, more and more patients and nursing staff own mobile phones that support Bluetooth Low Energy. Furthermore, deploying iBeacons that could be attached to trolleys or stretchers is quite easy and inexpensive. Therefore, it is fairly easy to implement an application that triggers an alert when a staff member moves away from their medical furniture.

However, today, no simple solution takes advantage of the fact that smartphones can be located through Wi-Fi and iBeacons attached to medical furniture with smartphones. This would provide an inaccurate but inexpensive, easy-to-deploy solution for locating furniture outside highly equipped rooms.

Based on this observation and the lack of existing solutions for combining different geolocation solutions, we have designed a new solution to solve this problem. Our solution tries to facilitate the integration of new location solutions. Its architecture is also designed to be distributed, support multiple users, and allow for multiple positioning levels.

The rest of this paper is organized as follows: Section II provides an overview of related research in geolocation systems that aim at combining several kinds of location technologies. Section III presents the proposed solution and its architecture. Section IV describes its technical implementation in detail. Finally, Section VI concludes the paper and provides directions for future work.

II. RELATED WORK

By location systems, we actually mean very different technologies that aim at tracking people or goods, providing services intended for end users or infrastructure owners, for real-time service or for analytics [5][6]. In terms of hardware, the measurements are taken using users smartphones, or through antennas or sensors that are specifically deployed for this purpose, such as RFID readers and Wi-Fi routers. The underlying technologies are Global Positioning System (GPS) [7], Wi-Fi, Bluetooth Low-Energy (BLE), Ultra-Wide Band (UWB) [8], Light-Fidelity (Li-Fi) [9], video, Near Field Communication (NFC), Quick Response codes (QR codes), smartphone gyroscopes, compasses, etc.

The located devices are passive and active tags or smartphones, except for a few solutions such as video processing, which does not require any device to locate people and uses biometrics. Smartphones can act both as sensors and tags.

Two main categories of algorithms stand out. The first one is based on the estimation of the distance between the object to locate and several points whose positions are known (e.g. trilateration). Depending on the technology, the distance may

 ISBN: 978-1-61208-421-3

be calculated from different metrics (Receive Signal Strength Indicator (RSSI), Time of Arrival (TOA), Time Difference of Arrival (TDOA), Angle of Arrival (AOA), etc.) and processed with basic algorithms such as trilateration or fingerprinting [2]. Because of the multipath propagation and non-line of sight, the results are quite poor. For these reasons, a lot of research is still underway to improve these techniques, such as the introduction of the Inertial Measurement Units and the use of Hiden Markov Model, grid-based filter, Kalman filters, particle filters, etc. [5].

The second category is based on the association of several physical measurements with a position. It requires a training/calibration phase (e.g. fingerprinting), which is quite heavy work since someone has go to the site and measure the different signals with different devices. Moreover, if the physical configuration of the site changes, this measurements are not relevant anymore. Research is being carried out on solutions for crowdsourcing these data in order to automate this task [10].

Hybrid technologies and signals of opportunities are promising solutions as they use every available source of information to improve accuracy and availability [11][12][13]. For example, they use smartphones 3-axis accelerometers and digital compasses to calculate the motion dynamics information and combine it with RSSI Wi-Fi positioning. These solutions also make use of signals from legacy systems such as radios, television sets, Wi-Fi and mobile phones to calculate positions within a few meters where GPS is not available.

All the aforementioned solutions are interesting technologies that aim at improving accuracy and robustness by focusing on the location of one object with one particular approach. Several of these proposals are mature enough and have already been deployed. Nevertheless, these solutions are developed in silos, and there is no seamless location service.

To overcome this technological limitation, several issues have to be addressed. One of them is the lack of a common solution for representing indoor/outdoor geographical data. To address this issue, the Open Geospatial Consortium (OGC) is working on IndoorGML [14] to model spatial information.

Another limitation is the lack of an architectural solution for combining different location technologies. The i-Locate project [15] focuses on providing a toolkit and a platform for using and managing several indoor/outdoor technologies such as Quuppa [16] and Galileo [17]. At the end of the project in 2016, the I-Locate *core localization service* should integrate those technologies and several services such as a location resolver, identity management or asset management. This open-source approach is very interesting, especially because it will be tested on site in several Italian hospitals. Nevertheless, based on the project specifications, the proposed solution is monolithic and heavy, as it is based on a 3-tier server architecture. This prevents the easy deployment of part of the data processing on mobile phones or other parts of the architecture, such as a gateway.

Moreover, even though the solution will make it possible to guide someone using different location technologies during a single journey thanks to its routing service, it will not provide any mechanism for combining technologies and making the most of all the available devices simultaneously.

Our work does not focus on improving existing geolocation techniques, but rather on efficient integration between different technologies. Compared to other projects with the same objectives we aim at providing a system with decentralized intelligence.

III. ARCHITECTURE

The whole architecture mainly relies on a *publish/subscribe broker*. The task of the broker is to receive messages and serve them to those who need them. The main idea is that antennas, locators and algorithms are *clients* of the *broker* (Figure 1). A client can subscribe to topics (channels) and publish messages on topics. When a client publishes a message on a topic, all the clients who have subscribed to that topic receive the message instantly. Sent messages can be marked as retained so they are kept on the *broker* and sent to *clients* when they subscribe to that *topic* later on.

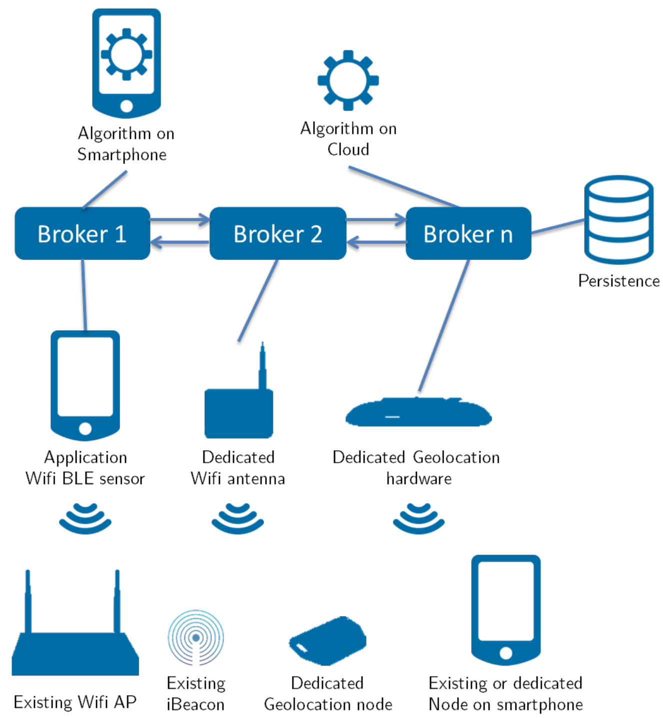


Figure 1. Architecture: all the blocks connected to the brokers are clients. The three clients under the brokers are called antennas.

With this architecture, we aim to decentralize everything as much as possible. Each sensor, algorithm or client can run on different devices as soon as they are connected to the broker. The broker can also be decentralized as will be explained later in Sections IV-A and VI.

A. Identifiers

To track nodes or users using multiple technologies, we need a way of doing so using these different technologies. For instance, a node may be identified by its Bluetooth and Wi-Fi Medium Access Control (MAC) addresses. To know that the node identified via Bluetooth and the node identified via Wi-Fi are the same, we need to declare that both identifiers are linked. To do so, a node or user must declare identifiers for all

 ISBN: 978-1-61208-421-3

the different technologies that it uses. To do so a message has to be sent on different topics. The message is always the same and is a dictionary of *Technology: Identifier* plus a *timestamp*:

```
{"BLE":"aa:bb:cc:dd:ee:ff",
"WIFI":"11:22:33:44:44:66",
"timestamp":1425384812592}
```

This message is sent to the following topics:

```
Identifier/BLE/aa:bb:cc:dd:ee:ff
Identifier/WIFI/11:22:33:44:55:66
```

These messages are retained on the *broker* so they only need to be sent when the device connects for the first time.

B. Measures

The measurements can be taken by any device that can sense its environment and communicate with the broker. Measurements can be Wi-Fi probe requests, iBeacons sensing, QR code or NFC tag reading, inertial measurements and so on. Measurements are sent in JSON format on a specific topic. For instance, the device with Wi-Fi Mac address `00:11:22:33:44:55` that receives a beacon packet from `aa:bb:cc:dd:ee:ff` with a signal strength of -59 will send the message:

```
{"timestamp":1422274308,"rssi":-58}
```

on the topic :

```
Measure/00:11:22:33:44:55/WIFI/aa:bb:cc:dd:ee:ff
```

C. Location requests

When a client wants to know its own location or that of another device (e.g. the location of an iBeacon), it has to request it. This way, if a device gets sensed by some sensors but nobody asks for its location, no algorithms run uselessly. Location algorithms are instantiated on demand. To request a location, a device needs to publish a retained message on the topic:

```
Location/Request/<requester>/<located>
```

where `<requester>` is the identifier of the client that requests the position, and `<located>` is the id of the device to locate. While the message exists on the broker, algorithms will try to compute the located position. When the locator does not need the `<located>` position anymore, it removes the message from the *broker* by sending a null message on the same topic to inform the algorithms that they can stop locating the device.

D. Referential

A referential is an abstract area associated with a position. A referential can contain an image of a map associated with an origin on the map and vectors. This way, there is no need to map an area to the World Geodesic System, and new areas can easily be added.

E. Algorithms

An algorithm is associated with an algorithm helper in charge of listening for location requests that involve algorithm. The algorithm helper can be in charge of several different algorithms. When the algorithm helper receives a location request, it verifies that the algorithms that it handles are capable of answering the request, and it eventually instantiates one or several algorithms to handle the location request.

When an algorithm is running, it listens for a certain type of measurements for the located device. The measurements are often associated with another device than the `<located>` one, which may be the locator or an antenna. To compute the located position, the algorithm may need the positions of locators or antennas. The algorithm will then request the location of that device. If the position is fixed, it may have been set as retained in the broker, so the answer is known instantly. Otherwise, another algorithm may answer the request. This mechanism makes multilayer positioning possible.

The position computed by the algorithm is sent on the topic: `Location/<located>/<algo_id>`.

F. Fusion

Several algorithms may compute the position of one device. They will post the position that they computed on the subtopic according to their names. In order for a final client to know its location, the results from the different algorithms have to be combined. A Fusion algorithm is in charge of combining the different positions computed by the different algorithms, and posting the merged position on the topic: `Location/<located>`. Like other algorithms, Fusion algorithms may be executed on a smartphone, server or any device that can connect to the broker.

G. History

In addition to the retained messages that the broker keeps, a history of measurements, locations and identifiers is kept. A specific client listens for all messages on the *broker* and stores them into a database. For now, this history is not used, but instead of waiting for measurements, *Algorithms* can request them in the history to speed up cold starts. The history may also be used to know the position of a device at a given time in the past.

IV. IMPLEMENTATION

To test our architecture, we have deployed several components: a Message Queue Telemetry Transport (MQTT) Broker, an Android mobile application, a modified wireless router and a cloud application server.

A. MQTT Broker

Our implementation relies on existing technologies and some that were specifically developed for the project. We use the open-source *Mosquitto* MQTT broker [18] for the *publish/subscribe* service. Several instances of the broker can be deployed on different servers to balance the load. Testing brokers capability is one of our ongoing tasks.

B. Android Mobile application

An Android mobile application has been written as a mobile client. Its purpose is to connect to the Broker and send iBeacon measurements. It also searches for Wi-Fi networks every 5 seconds in order to feed data into the wireless router.

C. Wireless router

Three TP-LINK MR3020 routers have been modified to listen for Wi-Fi Beacon frames and send them to a Broker. The *OpenWRT* [19] firmware has been installed on these routers and a custom application written in C connects to a Broker and is listening for 802.11 beacon frames. The application uses *libmosquitto* [18] and *libpcap* [20].

D. Cloud server application

Some algorithms and a fusion algorithm have been implemented on the server side. They are implemented in Java and connect to the MQTT broker. The algorithms implemented are trilateration over Wi-Fi, trilateration over Bluetooth, and a simple fusion algorithm that computes the barycenters of given positions.

E. In-browser viewer

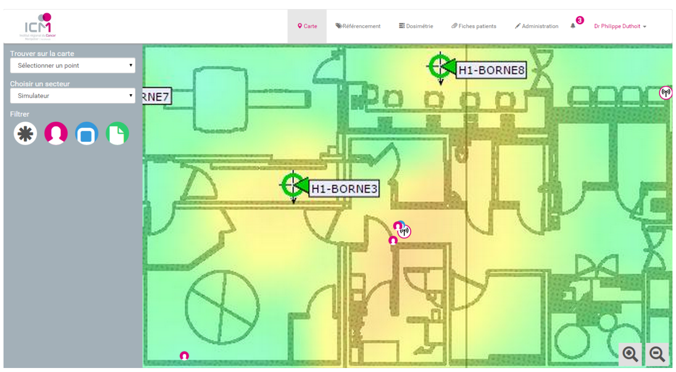


Figure 2. Web based viewer

The *mosquitto* MQTT broker can handle *websockets* connections; therefore, a web browser can be used as viewer. A web page has been developed that connects to the broker (Figure 2). The user selects an object to locate. Then the viewer displays the referential associated with the position and draws the object on it. A referential can also be selected so it is displayed, as well as all the objects located on this referential.

V. Results

The working architecture have been compared with the location service of an Android smartphone using Google Maps Android Application [21] and Quuppa [16] tags and antennas.

The test scenario is to start from a parking lot outside a building, enter the building and then enter a room equipped with Quuppa antennas. The building is also equipped with three Wireless routers described in Section IV-C. The test is made by a person holding an Android smartphone and a Quuppa tag. The smartphone have our application running in background and the Google Maps application in foreground. High precision location and Wifi is activated on the device.

On the Google Maps application the localization is accurate with an error of 5 meters when outside. As soon as the smartphone enter the building, the indicated position is off by up to 25 meters and the device is always located outside the building. As the device move in the building the indicated position barely move while staying outside the building, increasing the error of positioning.

On the Quuppa positioning system, the position reported by the system is not available until the tag is in the room equipped by the Quuppa antennas. In that room the error made on positioning the tag is less than 0.20 m.

Finally, using our architecture to locate the smartphone, we measure similar precision than using the smartphone location service outdoor while increasing it indoor. Indeed when the smartphone is in the range of the wireless routers the system computes an estimation of the position with an error of less than 10 m.

These early results show that the platform is able provide a continuous location service while using several technologies indoor and outdoor. We show that we can achieve continuity of positioning using a unique system while keeping a precision in the same order of existing systems used separately. For now Quuppa has not been integrated in the architecture but it is planed as a future work. It would lead to a positioning system combining strength of all positioning system we have at our disposal while providing ability to switch from one system to another seamlessly.

VI. Conclusion and Future Works

Future work will focus on features, security and privacy, third-party integration and performance. Planned features are: more compact messages, the ability for referentials to be positioned in other referentials, and running a MQTT broker on mobile phones. With a MQTT broker on mobile phones, if an algorithm is also running on the mobile phone, there will be no need to send messages to a broker located on a remote server. Actually, the brokers will communicate with each other and only transfer the messages that need to be transmitted.

Security and privacy will also be addressed using MQTT brokers capabilities. The objective is to fully isolate data so unauthorized users cannot see other users data and locations.

Finally, we plan to integrate the Quuppa [16] and Ubudu [22] geolocation technologies as black-box algorithms.

Through this article, we have proved the concept of a decentralized geolocation architecture that support multiple technologies and layers. This work will lay the foundations for future experiments with multiple technologies and devices, and will be a business enabler for our customers.

 ISBN: 978-1-61208-421-3

REFERENCES

[1] M. Kouhne and J. Sieck, “Location-based services with iBeacon technology,” in Artificial Intelligence, Modelling and Simulation (AIMS), 2014 2nd International Conference on, Nov 2014, pp. 315–321.

[2] Z. Yang, C. Wu, and Y. Liu, “Locating in fingerprint space: wireless indoor localization with little human intervention,” in Proceedings of the 18th annual international conference on Mobile computing and networking. ACM, 2012, pp. 269–280.

[3] E. A. Fry and L. A. Lenert, “Mascal: Rfid tracking of patients, staff and equipment to enhance hospital response to mass casualty events,” in AMIA Annual Symposium Proceedings, vol. 2005. American Medical Informatics Association, 2005, p. 261.

[4] “EKAHAU,” (Accessed: 2015.07.07). [Online]. Available: http://www.polestar.eu/en/use-cases/shopping-malls-large-retailers.html

[5] Y. Gu, A. Lo, and I. Niemegeers, “A survey of indoor positioning systems for wireless personal networks,” IEEE Communications Surveys Tutorials, vol. 11, no. 1, 2009, pp. 13–32.

[6] Z. Farid, R. Nordin, and M. Ismail, “Recent Advances in Wireless Indoor Localization Techniques and System,” Journal of Computer Networks and Communications, vol. 2013, Sep. 2013, p. e185138. [Online]. Available: http://www.hindawi.com/journals/jcnc/2013/185138/abs/

[7] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, Global Positioning System: Theory and Practice. Springer Science & Business Media, Apr. 2013.

[8] C. E. Baumann, L. Carin, and A. P. Stone, Ultra-Wideband, Short-Pulse Electromagnetics 3. Springer Science & Business Media, Apr. 2013.

[9] A. Lavanya, S. Umar, K. A. Kumar, and A. Kiranmai, “Light Fidelity: The New Wireless Communication Technology,” Journal of Mobile Computing, Communications & Mobile Networks, vol. 1, no. 1, May 2014, pp. 20–23.

[10] H. Shin, Y. Chon, Y. Kim, and H. Cha, “A Participatory Service Platform for Indoor Location-Based Services,” IEEE Pervasive Computing, vol. 14, no. 1, Jan. 2015, pp. 62–69.

[11] J. Liu, R. Chen, L. Pei, R. Guinness, and H. Kuusniemi, “A Hybrid Smartphone Indoor Positioning Solution for Mobile LBS,” Sensors, vol. 12, no. 12, Dec. 2012, pp. 17 208–17 233. [Online]. Available: http://www.mdpi.com/1424-8220/12/12/17208

[12] W.-L. Li, R. Iltis, and M. Win, “A smartphone localization algorithm using RSSI and inertial sensor measurement fusion,” in 2013 IEEE Global Communications Conference (GLOBECOM), Dec. 2013, pp. 3335–3340.

[13] C. E. Galván-Tejada, J. C. Carrasco-Jiménez, and R. F. Brena, “Bluetooth-WiFi based Combined Positioning Algorithm, Implementation and Experimental Evaluation,” Procedia Technology, vol. 7, 2013, pp. 37–45. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212017313000066

[14] Y.-J. Kim, H.-Y. Kang, and J. Lee, “Development of Indoor Spatial Data Model Using CityGML ADE,” ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 2, Aug. 2013, pp. 41–45.

[15] “i-locate,” (Accessed: 2015.05.04). [Online]. Available: http://www.i-locate.eu

[16] “Quuppa - Do More With Location,” 2015, (Accessed: 2015.05.04). [Online]. Available: http://quuppa.com

[17] “GALILEO - CNES,” (Accessed: 2015.05.04). [Online]. Available: http://www.cnes.fr/web/CNES-en/1507-galileo.php

[18] “Mosquitto,” 2015, (Accessed: 2015.05.04). [Online]. Available: http://mosquitto.org

[19] “OpenWrt,” 2015, (Accessed: 2015.05.04). [Online]. Available: https://openwrt.org

[20] “Tcpdump/Libpcap public repository,” (Accessed: 2015.05.04). [Online]. Available: http://www.tcpdump.org

[21] “Google Maps Android Application,” (Accessed: 2015.07.07). [Online]. Available: http://www.google.com/maps/about/

[22] “Ubudu | Mobile Proximity Solutions for Businesses,” (Accessed: 2015.05.04). [Online]. Available: http://www.ubudu.com/

 ISBN: 978-1-61208-421-3

An "Internet of Things" Vision of the Flood Monitoring Problem

Massimo Ancona*, Andrea Dellacasa*, Giorgio Delzanno*, Andrea La Camera* and Ivano Rellini†
*DIBRIS, University of Genoa
Genoa, Italy
Email: {massimo.ancona, giorgio.delzanno, andrea.lacamera}@unige.it, andrea@dellacasa.org
†DISTAV, University of Genoa
Genoa, Italy
Email: rellini.ivano@dipteris.unige.it

***Abstract*—River flood monitoring is a complex problem of maximum social relevance in densely inhabited areas. Flash floods are becoming more and more dangerous every year due to an increase of rapid and extreme rainfalls events induced by climate changes, particularly in Genoa: the rivers flooded twice in twenty days in two different parts of the town. The complexity of the problem originates from the diversity of the territory involved in the monitoring process in regions like Liguria in Italy: from relatively far mountains (or plain regions) generally scarcely populated, to densely populated urban areas traversed by streams, often flowing underground. Environmental monitoring is classified at the 14th position among the Top 50 "Internet of Things" (IoT) applications for a Smarter World. IoT is an emerging paradigm that combines the main features of Cloud Computing (pervasive capability of storage and computation via Internet) with wireless sensor networks that provide cooperative use of distributed sensors. In addition to these features, IoT often provides interfaces for data streaming management in real-time, back end for data analysis and visualization. A IoT approach to the considered problem seems almost necessary in order to improve the integration level of the data, sensors and applications and provide software tools for cooperation of heterogeneous groups of end-users involving institutions with different level of responsibilities, knowledge and capabilities (from sensor experts to non-expert citizens). In the present paper, we redesign previous experiences based on intelligent wireless sensors network (WSN) in terms of IoT with the aim of improving its reliability and efficiency and especially of rendering the problem widely scalable.**

***Keywords*–Internet of Things; Communication networks; Stream; Flood; Sensor systems.**

I. Introduction

Flood monitoring is a particularly challenging application for Internet of Things (IoT). In fact, it offers a complex scenario for the variety and number of sensors involved, their location and relative communication problems. The type of sensors involved in the process and the corresponding type of installation depend on the kind of collected data and on their geo-localization (i.e., urban areas, where powering and communications are relatively simple, or in remote and difficult to access mountainous or country locations). The kind of data collected ranges from rain monitoring to river gauging with several parameters to be monitored and compared. In the case of rivers, the problem depends on their size and dimension and geography of the region where they flow, if they are small creeks or wide rivers, if they flow in a steep or flat area, in open air or are channeled underground, etc. From this point of view, we already activated different collaborations and definitions of common goals with public administrations involved in the management of the experimental areas. To this aim, we designed a general hardware and software IoT infrastructure and architecture applicable to the environmental problem mentioned above, but extensible to the more general problem of monitoring the environment in densely inhabited areas.

Our research will be an element of great importance to train specific risk management and to deliver elements of innovation and encouragement for the definition of land management strategies both on the local and regional scale. Moreover, this research will help to provide knowledge and tools for effective decision making and public engagement. In particular, we detail the sensor classes (their design for the new ones), their communication mechanisms and associated software services as components of a general IoT infrastructure. The aim is to monitor either rainfalls, river discharge and their temporal correlation in order to obtain early alarming information. In our IoT approach, all collected data will be continuously transmitted, through the Internet communication infrastructure, to software components designed to compute the stream-flow and to quantify the spatial distribution of flood risk for each controlled watershed. The computed risks, together with data coming from other sources (barometric and river discharge sensors, cameras operators of public organizations, emergency agencies, private citizens), will be examined by a diagnostic decision system implementing a risk-alert scheduling strategy, able to diagnose the health state of the controlled environment and to define specialized alarm levels for each potentially interested area. Finally, the computed risks will be used for specializing alerting messages, to be sent to all citizens (ubiquity) present in each selected area only (alerting locality).

The interaction between instrumental data with other sources of information, including people, is another objective of our research. A connected aspect of the problem is the complexity of alarming organization and broadcast mechanism deriving from the inherent uncertain and aleatory nature of the estimates that strongly affect the management of related information [1][2], while the interaction of instruments with other sources of information, including people, is another objective of our research. In this paper, we present an approach based on intelligent WSN [3] in terms of IoT. Our research is intended for improving its reliability and efficiency and especially for rendering the problem widely scalable, from small towns to medium and large cities. In Section 2, we describe the different kind of sensors, and in Section 3 we present sub-1GHz network and standardization. Section 4 is dedicated to software needed to run the system. We summarize the conclusions in Section 5.

 ISBN: 978-1-61208-421-3

II. Sensor evolution towards an ultra-low-power architecture and M2M connection to Sub-1 GHz IoT networks

The early warning information is represented by rainfalls estimation and weather forecasting in a given stream basin. Flash-flood monitoring requires rainfall control at a dense spatial grid (1 km or finer) and frequent time-scale intervals (15-30 min, and even less in urban areas) [4]. Our analysis is based on the above assumption that impacts the kind of sensors considered, their communication mechanisms and management software. The kind of sensors for a near real-time control should be maintenance free, i. e., free from mechanical moving parts. Moreover, they should communicate via the Internet via a machine to machine (M2M) technology, which is typical in the IoT implementation, adopting an ultra-low-power management system, in order to operate for months from remote areas, supplied by small batteries. Note that the proposed IoT architecture can be extended (by replacing the specific sensors) to other environmental problems, e.g. for controlling landslides, another critical aspect of a territory. Legacy sensors are not ready for the IoT revolutions, because of power consumption, lack of wireless communications, and costs.

A. Rain gauges

If we consider a spatial sensor grid of 1 km the number of gauges installed grows as n^2 requiring 100 gauges for a grid of 10 km^2, a very high number raising cost and network complexity. This approach requires very cheap and reliable gauges. Cost reduction suggests to consider non mechanical (i. e., with no moving parts) rain gauges to reduce the need of ordinary maintenance of devices often installed in remote areas. The market is offering several kinds of devices: optical, ultrasound and based on other measuring principles, but generally they are not designed for ultra-low-power applications and Internet access in an IoT M2M connection. Thus, we have to shield them with an ultra-low-power microprocessor unit (MPU: a computer processing unit in a single integrated circuit), a technique used by Arduino and other MPU-based devices [5] for interfacing with peripheral accessories. Optical rain gauges (ORG) are able to provide ultra-fast real-time data about precipitation rate, with minimal maintenance, high reliability and sensitivity, We found both high precision (3%), power hungry (12V 1A, or direct AC power line connection), costly devices (5-7 k$), like the Osi 815-DS [6] and the All Weather Inc. 6030 [7]; and low power, low cost devices (50$), like the Hydreon RG-11 [8]. In the last case, errors usually are small (around 6% in mean) but not granted, and in some (rare) cases they could reach 30%. There exist also low cost, accurate ultrasound and electronic gauges; although they empty themselves automatically, they do not require adjustments and do not use moving parts. In conclusion, we can find on the market several adequate rain gauges, but we have to shield them with an ultra-low-power MPUs board (see Figure 1) for long-time, low power (and long distance) applications, when gauges are installed in remote areas. ORG are the more promising devices for several reasons: they are fast and maintenance free, they can measure true rain rate and no other gauge comes close to their performance and precision. Nowadays one of the main applications of low cost ORG is the automatic control of car wipers: for marketing reasons this devices will be rapidly improved in the next few years. Almost all devices communicate via a serial RS232 or 485 interface. The RG-11 drains 15mA nominal at 12V DC input, but can be operated also in a slightly less sensitive micro-power mode that can be set which should allow operation from a 9V battery at 1.5mA current draw. It does not guaranties a strict accuracy value below 28-36% of a tipping bucket, but much of the time it will read very close to it. The RG-11 is great for qualitative measurements versus quantitative, i.e., you will know when there is a heavy rain rather than a light mist but not an exact amount of rain, in other words, it is well suitable for our scope. By installing a large number of low cost devices alternated with few high precision gauges we may grant, through statistical corrections, a good precision at an acceptable cost.

B. River Gauges

The river gauging problem is more complex than the rain gauging for its intrinsic complexity: most of the short stream crossing Liguria and other similar hilly regions have an almost binary behavior. Their flow is extremely small or null for a large part of the year, while suddenly it becomes very impetuous for short time periods during October and November rain storms. This behaviour often generates disastrous flash-floods. During rain storms periods, floods eradicate trees, rocks and instruments working immersed in the water flow and they tend to modify their bed often in a remarkable way by transporting stones and trees that could obstruct their flow in a stable way. Generally, the flood phenomenon (flash-flood) lasts for few hours and produces incredible disasters. In these conditions, the river gauging problem represents one of the most critical part of our research and is, in large part, still an open problem. The available gauges on the market present several restrictions on their applicability. Also in this case, we limit our attention to non mechanical (i.e., with no moving parts) gauges for maintenance purposes. Moreover, we restrict our attention to devices not operating immersed in the stream, because their expected survival should be very short. There are some gauges that implement, in a single device like the Marsh-McBirney Flo-Dar [9]area/velocity flow meter, both measurements of water level and of flow velocity in order to compute stream flow. Their present limit is that they should be mounted above the flow at 1-3 m distance (optionally extensible to 6 m, which is a safe limit for our cases). New approaches based on particle imaging velocimetry (PIV) that give an accurate flow measurement, are appearing in literature. A promising approach based on PIV gauges is proposed in [10] where a digital camera installed on a roof of a building at 14 m above the river level and with an incident angle of 60^o provides a maximum error of 38.8% while its mean error is only the 5%. We plan to extend a PIV gauge with a small Light Detection And Ranging (LIDAR) (or two micro-cameras) for computing distance and reconstructing a 2D section of the bed of the controlled stream during the long no-raining periods when the stream is near empty. As an alternative, we could use a drone for reconstructing the complete section of the stream in all its extension during a single or few drone flies, but this approach requires operator's intervention.

C. Integrating gauges in an IoT hardware communication architecture

Most of the gauges offering the best precision are often power hungry devices, a defect limiting their installation in remote locations where power and network are not present. Indeed a precise information derived by a well distributed set of gauges could have a higher impact on the whole acquisition system (following a well geo-located gauge network) [4]. In this case, the IoT technology offers the best solution for ultra-low-power energy use and communication continuity in locations uncovered by power and communication lines. The IoT approach, while minimizing power consume, offers the advantage of making gauges able to communicate each-other via Internet in M2M communication mechanism satisfying an anything-anywhere connectivity framework independently from the availability of power systems and network access. The access to the Internet rises security problems that at an early research stage have a lower priority than the requirements of minimum power consumption and communication continuity. Because most of the rain and river gauges available today are not designed for power saving, the IoT based shielding approach, shown in Figure 1, solves this problem. A sensor is activated only for the minimum time needed for making a measurement, while the ULPM (Ultra-Low Power Micro-controller) together with other power control circuits manages the sleeping versus active state between two successive measurements.

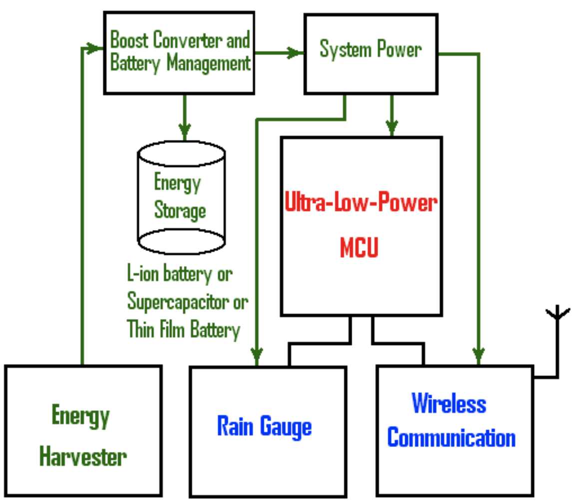


Figure 1. Gauge shielding architecture

In Figure 1, the sensor shielding architecture is composed by the sensor itself, a ULPM, a buck converter (e.g., tps62740 or lt3757a) a boost charger (e.g., bq25504) and a ultra-low-power wireless communication sub-system (e.g., Texas Instruments CC3100 for Wi-Fi connection to the Internet).

III. The Sub-1GHz Network and Relative New Standards

For covering a wide variety of cases, in different parts of the environment with different communication requirements, the network must adopt different wireless communication standards to support M2M connections between two connected devices, without the assistance of a human. Here, we mention the more likely to be used: Wi-Fi technology, based on the IEEE 802.11, today represents the widest wireless protocol adopted for short distances (100 m). Wi-Fi is so intimately integrated with the TCP/IP that the Wi-Fi term implicitly mean that they are also using a TCP/IP for Internet connectivity. Finally, the new *Weightless* (expressly designed for M2M communications), the IEEE 802.11ah and the IEEE 802.11af standards, all operating in the Sub-1GHz (together with many proprietary radio systems, and other well established standards yet working in the Sub-1GHz bands) are on the way [11]. They are capable to transmit over several km within a simple point-to-point or star topology all performing with an extremely low power consumption. To connect to the IoT, Sub-1 GHz systems require an application-layer Internet gateway that, in our case, is provided by the ultra-low-power MPU shielding every sensor and including the communication device itself (e.g., CC3100 or CC430).

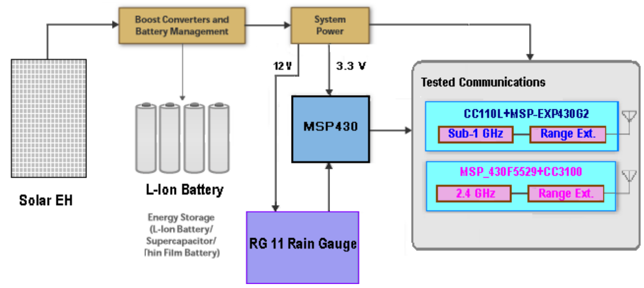


Figure 2. Implemented sensor node configurations

Figure 2 illustrates two configurations we have implemented for low power remote IoT/M2M connections. The CC3100 supports Wi-Fi short range connection, while the TI CC110L implements a long-range Sub-1GHz M2M wireless connection providing a ultra-low-power several km (up to 10-25) connectivity with the CC1120/CC1200 [12]. The Sub-1GHz frequency range is very suitable for rural regions [11].

IV. Software Infrastructure

Software requirements for IoT sensors node networks presents several challenging problems [13][14]. The system is widely distributed and able to orchestrate a wide variety of smart sensors, supported by energy harvesting mechanisms and managed by ULP-MPUs. The system follows an event-driven architecture centered on near-real-time asynchronous communication and control mechanisms [15]. The sensor shielding approach suggests a 2-tear structured network to the more common 3(or n)-tear layered architecture. Fault-tolerance is another fundamental requirement, because controlling a single event requires continuous and autonomous inter-device communication. The global interaction cannot be interrupted and must remain operational especially during stormy weather days, when it is required to be continuously operative. Collected data should be correlated in real-time for computing the lag time that is the time occurring between a rainfall peak and the corresponding expected discharge peak.

A. *IoT Platforms*

The sensor components are designed to be integrated in a IoT software platform built for integrating different types of services, sensors and data that could be used for improving the quality of social and environmental services of urban areas. An ideal IoT platform for our application domain should provide the following components:

 ISBN: 978-1-61208-421-3

- A device management component to handle registration of new devices, assignment of unique identifiers, format data, etc.
- Sensor services to provide interfaces to interconnect in a secure way heterogeneous information sources.
- Storage services to persistently store data.
- Analytic services to provide both predefined and customizable procedure to elaborate stored and real time data streams.
- Visualization services to disseminate collected data using different formats like visual diagrams, reports, graphs, etc. For the considered type of application, it is also important to consider georeferenced data dissemination of alarms and notifications.
- Application and user management to handle in a secure way registrations of users and of new applications that extend the functionality of the system.

For instance, alarms used in flood monitoring require real-time processing of data and rapid responses to public organizations. Furthermore, they require tight integration of data coming from very different sources (from sensors to sms sent by citizens via crowd sourcing applications). Another important aspect is the possibility of extending the set of available services with new components that could exploit the data collected by the system. Last but not least, there is the problem of taking the alarming decision based on flood forecasting and weather predictions methods [16].

Existing platform like Axeda [17], Thingworx [18], Thingsquare [19], Eclipse M2M [20], and Xively [21] provide powerful tools for setting up complex applications that combine data integration, analysis, and visualization. For instance, Thingworx provides a composer tool that helps the designer to set up a dashboard with graphical widgets to visualize the results of analysis of data coming from external services. Thingworx marketplace can be used to download and install additional packages, e.g., Google Maps widgets, device drivers, and database management system (DBMS) connection libraries. Platforms like Kaa [22], Kinoma [23], M2MLabs [24], Arduino provide similar features in an open source environment.

V. Conclusions and Future Work

Flash flood alarming systems require a dense network of rain gauges for monitoring intense local rain storms both to ensure its survival in case of extreme weather and to have a more accurate collection of data. Those data have to be interpreted by means of empirical and formal models by correlating in real time the river level and the flow intensity for early flood forecasting and consequent anticipated alarming. The number of required sensors, their communication mechanisms and reliability requirements show that a IoT/M2M approach is able to resolve the problem, even if today commercially available gauges are not designed with the specific objective of flash floods control. Their main limits are high power consumption, inadequate communication mechanisms and costs. However, by supporting gauges with a ultra-low-power MPU minimizing gauges operational time, an M2M network with minimal power requirements can be implemented. Moreover, our research shows that new rain gauges have to be designed with the specific objective of flood control and not for general purpose applications, like agriculture. Next steps in this project will be: a deeper testing phase of the network of low cost sensors both in simulated and real-life scenarios, and an implementation of the software behaviour in Thingworx or other platforms.

References

[1] H. McMillan, J. Freer, F. Pappenberger, T. Krueger, and M. Clark, "Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions," Hydrological Processes, vol. 24, no. 10, 2010, pp. 1270–1284.

[2] F. Pappenberger et al., "Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication," Hydrological Processes, vol. 27, no. 1, 2013, pp. 132–146.

[3] M. Ancona et al., "On the design of an intelligent sensor network for flash flood monitoring, diagnosis and management in urban areas position paper," Procedia Computer Science, vol. 32, 2014, pp. 941–946.

[4] M. Borga, E. Gaume, J. D. Creutin, and L. Marchi, "Surveying flash floods: gauging the ungauged extremes," Hydrological processes, vol. 22, no. 18, 2008, p. 3883.

[5] E. Gertz and P. Di Justo, Environmental Monitoring with Arduino. O'Reilly, 2013.

[6] "Org-815 - optical scientific," http://www.opticalscientific.com/pdf/press_releases/ORG815_PR%20120529.pdf, accessed: 2015-06-09.

[7] "6030 optical rain gauge - all weather inc," http://www.allweatherinc.com/wp-content/uploads/6030-0411.pdf, accessed: 2015-06-09.

[8] "Rg-11 - hydreon," http://www.rainsensors.com/docs/rg-11_instructions.pdf, accessed: 2015-06-09.

[9] "Flo-dar av sensor," http://www.hachflow.com/flow-sensors/flo-dar.cfm, accessed: 2015-06-09.

[10] J. Creutin, M. Muste, A. Bradley, S. Kim, and A. Kruger, "River gauging using piv techniques: a proof of concept experiment on the iowa river," Journal of Hydrology, vol. 277, no. 3, 2003, pp. 182–194.

[11] S. Ajah, A. Al-Sherbaz, S. J. Turner, and P. Picton, "Sub 1ghz m2m communications standardization: The advancement in white space utilization for enhancing the energy efficiency," in PGNET Proceedings of the 15th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting 2014, M. Merabti, O. Abuelmaatti, and C. Oliver, Eds. Liverpool, UK: Liverpool John Moores University, June 2014.

[12] "Long-range rf communication: Why narrowband is the de facto standard," http://www.ti.com/lit/wp/swry006/swry006.pdf, accessed: 2015-06-09.

[13] D. Hughes et al., "An experiment with reflective middleware to support grid-based flood monitoring," Concurrency and Computation: Practice and Experience, vol. 20, no. 11, 2008, pp. 1303–1316.

[14] P. J. Smith et al., "Towards the provision of site specific flood warnings using wireless sensor networks," Meteorological Applications, vol. 16, no. 1, 2009, pp. 57–64.

[15] S. Gusmeroli, S. Piccione, and D. Rotondi, "Iot@ work automation middleware system design and architecture," in Emerging Technologies & Factory Automation (ETFA), 2012 IEEE 17th Conference on. IEEE, 2012, pp. 1–8.

[16] "Reliable, low power wireless sensor networks for the internet of things: Making wireless sensors as accessible as web servers," http://cds.linear.com/docs/en/white-paper/wp003.pdf, accessed: 2015-06-09.

[17] "Axeda," http://it.ptc.com/axeda, accessed: 2015-06-09.

[18] "Thingworx," http://www.thingworx.com/, accessed: 2015-06-09.

[19] "Thingsquare," http://www.thingsquare.com/, accessed: 2015-06-09.

[20] "Eclipse m2m," http://iot.eclipse.org/, accessed: 2015-06-09.

[21] "Xively," https://xively.com/, accessed: 2015-06-09.

[22] "Kaa," http://www.kaaproject.org/, accessed: 2015-06-09.

[23] "Kinoma," http://kinoma.com/, accessed: 2015-06-09.

[24] "M2m labs," http://www.m2mlabs.com/, accessed: 2015-06-09.

A Smartwatch-Based System for Audio-Based Monitoring of Dietary Habits

Haik Kalantarian, Majid Sarrafzadeh
Wireless Health Institute, Department of Computer Science
University of California, Los Angeles
Email: {kalantarian, majid}@cs.ucla.edu

***Abstract*—In recent years, smartwatches have emerged as a viable platform for a variety of medical and health-related applications. In addition to the benefits of a stable hardware platform, these devices have a significant advantage over other wrist-worn devices, in that user acceptance of watches are much higher than other custom hardware solutions. In this paper, we describe the development of an Android application on a Samsung smartwatch device for evaluating eating habits using a microphone and various signal processing techniques. Though other works on acoustic monitoring of food habits have been conducted, the varied arm movement during eating creates a unique set of challenges that our work attempts to address. Evaluation results confirm the efficacy of our technique; classification was performed between apple and potato chip bites, water swallows, and talking, with an F-Measure of 94.5% based on 200 collected samples. *Index Terms*—smartwatch; nutrition; spectrogram**

I. Introduction

There is little doubt that obesity is associated with various negative health outcomes such as an increased risk for stroke, diabetes, various cancers, heart disease, and other conditions. In 2008, medical costs associated with obesity were estimated to exceed $147 billion, with over one-third of adults in the United States estimated to be obese [1]. The two major contributors to weight gain are an inactive lifestyle and poor diet. Though the former has been addressed by many wearable devices in recent years both in research and the consumer electronics field, few works exist on automatic detection of dietary habits in an inconspicuous form-factor [2][3][4]. Instead, characterization of an individual's eating habits is possible through manual record keeping such as food diaries, 24-hour recalls, and food frequency questionnaires. However, these approaches suffer from low accuracy, high user burden, and low rates of long-term compliance. Wireless health-monitoring technologies have the potential to promote healthy behavior and address the ultimate goal of enabling better lifestyle choices.

In recent years, several electronic devices have been proposed for monitoring dietary habits. However, most works attempt to characterize eating from patterns in chewing and swallow counts, and very few attempt to identify the nutritive properties of the foods themselves. Therefore, a fundamental question in the field of electronic food monitoring is the validity of chew and swallow counts as a heuristic for estimation of Caloric intake. A very recent work by Fontana et al. [5] addresses this issue by comparing several different techniques for estimation of Caloric intake: weighed food records (gold standard), diet diaries, and electronic sensor-based measurements of chews and swallows. The conclusion of this study was that chew and swallow counts were more closely affiliated to the gold standard measurement than self-reporting methods and photographic food records.

Fig. 1. A high level architecture of the proposed system is shown above. Many different forms of eating can be detected using a smartwatch, provided the appropriate hand is used and the watch is brought close enough to the mouth.

Many prior works address the problem of nutrition monitoring by processing audio signals associated with ingestion [6][7]. Typically, these systems use a throat microphone for recognizing deglutition (swallows), or using time-frequency decomposition techniques, such as Wavelet Packet Decomposition (WPD) or Spectrogram Analysis to extract distinctive features, and either classify between different food groups or recognize anomalies in swallow patterns. While many of these works are novel from a perspective of algorithmic techniques, they generally propose custom hardware solutions or bulky non-standard equipment which are of limited use outside of clinical environments.

Recently, smart-watches have emerged as a new platform that provide several promising applications such as wrist-worn activity monitoring, heart rate tracking, and even stress measurement. Watch usage is well established and has a high level of social acceptance, as confirmed not only by our personal studies but by their ubiquity in day-to-day life. Furthermore, the smart-watch platform provides many useful services that can collectively improve user adherence rates, rather than specialized devices with just one application that may fail to sustain a user's interest.

This paper explores the idea of tracking eating habits using a custom Android application on the smart-watch platform. Though identifying eating-related gestures using wrist-worn devices is a viable application of the watch, the focus of our work is to explore the idea of using audio to detect eating behavior based on bites, rather than swallows as other works have done. A high-level system architecture is presented in Figure 1. The first step is audio-based acquisition of eating-related sounds such as bites, acquired from the microphone integrated within the smartwatch. After data acquisition, the audio is processed using various classifiers to identify the sound and infer the associated activity. After synchronizing with cloud services, the user is provided with information about their recent eatings habits, and appropriate feedback when necessary.

The paper is structured as follows. In Section II, we present related work. In Section III, we describe the system architecture. Section IV presents an overview of the proposed algorithms for classification. Section V describes the experimental procedure, followed by results in Section VI. Section VII concludes the paper.

II. Related Work

Many works have proposed detecting food intake using static microphone placement, generally on the throat. For example, the work in [8] uses acoustic data acquired from a small microphone placed near the bottom of the throat. Their system is coupled with a strain gauge placed near the ear. Other works attempt to characterize and address swallow disorders in seniors, such as dysphagia [9].

In the work by Amft et al. in [10], authors analyze bite weight and classify food acoustically from an earpad-mounted sensor. In [7], the authors present a similar earpad-based sensor design to monitor chewing sounds. Food grouping analysis revealed three significant clusters of food: wet and loud, dry and loud, soft and quiet. An overall recognition accuracy of over 86.6% was achieved. A more recent study using support vector machines have been able to reach swallow detection accuracies of up to 84.7% in an in-lab setting [11]. These devices are mounted very high in the upper trachea, near the laryngopharynx.

In [12], Pler et al. proposed a system geared towards patients living in ambient assisted living conditions and used miniature electret microphones which were integrated into a hearing aid case, and placed in the ear canal. In [13], the authors are able to achieve a food detection accuracy of 79% using hidden Markov models based on data acquired from microphones in the ear canal.

III. System Architecture

Our proposed system does not require any custom hardware: the Android application runs on Samsung Galaxy Gear smartwatch running Android 4.2.1. This phone features an 800 MHz ARM-based processor, 512 MB of RAM, and a 320x320 pixel 1.6 inch display. The device also supports transfer of data using the Bluetooth LE protocol, and can be configured to access the Internet using Bluetooth tethering with compatible smartphones.

Data was recorded using the Samsung Galaxy Gear microphone in MPEG-4 Part 14 (m4a) format at a rate of 96 kbps, as prior research has shown that the spectral energy for many common foods is between 0-10 kHz, with highest amplitude ranges between 1 and 2 kHz for water [14][15].

IV. Algorithm Design

A. Frequency-Domain Evaluation: Liquids

We begin our algorithm analysis with the objective of detecting liquid ingestion using a smartwatch. Because we have a-priori knowledge about the kind of data we would like to identify, we could pre-process the recorded data before classification, as we describe in this section.

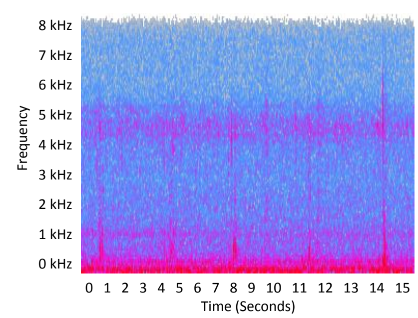


Fig. 2. A spectrogram of an audio clip consisting of five swallows, generated with a window of size 1024 samples. There is a visible change in the spectral density at points corresponding with swallows as shown above.

Figure 2 shows a spectrogram corresponding with an audio clip consisting of five water swallows acquired from the smartwatch. This spectrogram is generated with the Short-Time Fourier Transform, and shows changes in the frequency distribution over time [6]. Figure 3 shows a more detailed comparison between a brief interval of noise (1s) and a water swallow. Generally speaking, the data of interest is between 600 Hz and 1 kHz, as shown by the deviation between the signals at this time, and confirmed by the spectrogram shown in Figure 2. We conclude that analysis of this frequency range is critical for classification of liquid swallows. This observation is confirmed by Figure 4, which shows the transformation of an audio signal corresponding with ten swallows. The top waveform is the original, while the bottom is the post-processed filter output in which noise is substantially reduced. This is achieved by band-pass filtering the audio data with cutoffs of 600 Hz and 1 kHz and a rolloff of 48 dB- meaning

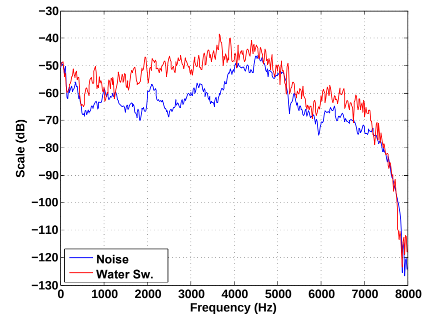


Fig. 3. Frequency distribution of a water swallow vs. silence (noise). This graph reveals that the frequency range between 500 Hz and 1000 Hz is the point of interest.

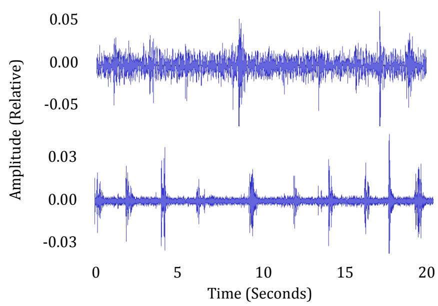


Fig. 4. Post-processing of the audio signal corresponding with water can dramatically improve signal-to-noise ratio. The top shows the original waveform. The bottom shows the waveform after a bandpass filter is applied.

the amplitude decreases by 48 dB for each octave outside the filter threshold.

While the resulting signal clearly shows the swallows, marked by pronounced peaks, this technique is not very generalizable to other foods besides water, because the data is pre-processed. In the case of the frequency distribution of a one second window around the initial bite of a potato chip, compared to an equal period of chewing, the amplitude of the bite signal is greater from 600 Hz to 4 kHz. However, the pattern is not as distinctive as for liquids, and may certainly vary between individuals with different eating styles. A more generalizable approach is described in the next subsection.

B. Generalizable Feature Extraction

Detection of eating habits is somewhat different than that of liquid consumption, as the smartwatch will not be near the throat during a swallow. Therefore, in these cases we attempt to identify when an individual bites into a food item rather than chewing. The smartwatch platform is particularly well suited for this application because the microphone will be nearest to the sound source during the times at which the signal is of interest. The proposed model must be flexible to identify biting and swallowing for many different foods and drinks, between individuals with varying eating styles.

openSMILE [17] is a feature extraction tool intended for producing large audio feature sets. This tool is capable of various audio signal processing operations such as applying

TABLE I. Partial List of openSMILE Speech Features from [16]

Speech-Related Features		
Signal Energy	Loudness	Mel/Bark/Octave Spectra
MFCC	PLP-CC	Pitch
Voice Quality	Formants	LPC
Line Spectral Pairs	Spectral Shape	CENS and CHROMA

TABLE II. Partial List of openSMILE Statistical Features from [16]

Speech-Related Features		
Means	Extremes	Moments
Segments	Samples	Peaks
Zero Crossings	Quadratic Regression	Percentiles
Duration	Onset	DCT Coefficient

window functions, fast-Fourier transforms,finite impulse response filterbanks, autocorrelation, and cepstrum. In addition to these techniques, openSMILE is capable of extracting various speech related features and statistical features. A partial list of extracted features is shown in Tables I and II, respectively. After data is collected from a variety of subjects eating several foods, classifiers can be used to identify strong features that are accurate predictors of swallows and bites for various foods, while reducing the dimensionality by eliminating redundant or weakly correlated features.

A microphone on a Smartwatch can either constantly record data, or be configured to record audio based on motion-based triggers indicative of eating-related gestures, in order to save battery life. The recorded audio is stored on a buffer in Smartwatch memory with storage for 4096 samples, corresponding with 0.25 seconds of data. Once the buffer is full, features are extracted using openSMILE (elaborated upon in subsequent sections), and the audio clip is classified divided into several distinct categories corresponding with the various foods the system has been trained to detect. A counter is incremented corresponding with the food type detected, which is necessary for long-term record keeping. In the event that eating behavior is detected, subsequent detection is disabled for a period of two seconds to prevent duplicate records caused by the same event. The algorithm is presented in Figure **??**, with $\beta = 4096$ samples and $\tau = 2$ seconds.

To minimize the overlap between neighboring segments for performance reasons, the last 50ms of buffer data are cleared after each classification activity, and classification resumes when the buffer is full once again (not shown).

V. Experimental Procedure

A. Data Collection for Recognition

A total of ten subjects were used for data collection, with ages ranging from 22 to 35 in order to develop a model for identifying swallows. The subjects included eight males and two females. Each subject was asked to eat the following foods, in order: three apple slices with at least two bites per slice, one 8 oz. glass of room-temperature water, and one bag of potato chips. The moments at which the food was bitten into (or swallowed as in the case of the water) were manually annotated.

```
RecordAudio(Buffer)
if Buffer.Utilization = β then
    d = Buffer[1:β];
    f = ExtractFeatures(D);
    s = {Water, Talk, Apple, Chips, Other};
    c = Classify(F,s);
Counter_c++;
if c ≠ Other then
    PauseRecording(τ)
```

Fig. 5. Simplified Classification Scheme

Data was manually extracted from the audio recordings at a later time. Regardless of the food or activity type, each sample was exactly 0.25 seconds in length, and the peak of the wave amplitude was not necessary centered in the window. In some cases, such as during the biting of an apple, one quarter of a second was not sufficient to capture the entire bite.

Subjects were also asked to read a brief passage from a Wikipedia article, with no particular instruction about the rate at which they should read. The data was then automatically split into 0.25 second audio fragments using an audio processing program. Therefore, some samples were collected between phrases, and were relatively silent.

B. Smartwatch Feedback: A Survey

Before the system development phase, we had several important questions about how individuals feel about smartwatches. As described previously, a wearable device must have both high accuracy, and high rates of user adherence for the subject to reach his or her intended goals. Furthermore, we proposed several questions about which hand a subject prefers to wear a watch. For example, our experimental evaluation requires that subjects wear a watch on the same hand with which they typically eat food such as chips or raise a glass of water.

An online survey was conducted with a total of 221 responses in which various questions were posed with respect to how individuals feel about wearing a smart-watch. The participants in the study were anonymous, but represented a diverse set of ages, cultures, and genders. The study was originally conducted on January 28th for an internal data collection on smartwatch usage applied to the domain of medication adherence, but we found the majority of the questions were also applicable to food-intake monitoring. The survey consisted of a total of 9 questions. Partial results and discussion can be found in Section VI.

VI. Results and Discussion

A. Audio Classification

Results for classification between apples, chips, water, and speaking are shown in Table IV based on 50 unprocessed samples collected from each of these foods, using the Random Forest classifier [18] with 6555 extracted features from each sample. The Random Forest classifier consisted of 100 trees, each constructed using 13 random features, and was validated using 10-fold cross validation. This particular classifier was chosen for its high accuracy in our experimentation- several other classifiers performed poorly in comparison. A total of 189 instances were classified correctly (94.5%) while the remaining 11 (5.5%) were classified incorrectly. The weighted average precision, recall, and F-Measure was 94.6%, 94.5%, and 94.5% respectively. Classification of water and speaking were particularly accurate, with only one incorrect classification. The majority of classification errors were between apples and potato chips.

B. Feature Extraction

From the 6555 extracted features, the Correlation Feature Selection (CFS) Subset Evaluator was used to evaluate 991,139 subsets of features. This subset evaluator considers both the individual predictive ability of features, as well as the redundancy between them, and found the merit of the best subset to be 0.948. The search was stale after 5 node expansions. In other words, the subset evaluator aggregates the best features linearly beginning with those that show the highest correlation, and terminates after five consecutive subsets show no improvement in classification accuracy.

The top ten features are listed in Table III. The first feature is the skewness of the logarithmic signal energy, in which skewness is defined as the asymmetry of the variable in comparison with a normal probability distribution [19]. More formally, skewness is defined in below, where μ_i is the ith central moment about the mean.

$$\gamma_1 = \frac{\mu_3}{{\mu_2}^{3/2}} \tag{1}$$

The ith moment M_i moment of a discrete function f(x) defined on an interval [a,b] can be generalized as:

$$M_i = \sum_{x=a}^{b} x^i f(x) \tag{2}$$

To calculate the moment about the mean for a probability density function, it is necessary to first calculate the mean *m*. The ith moment about the mean can be represented as:

$$\bar{M_s} = \frac{\sum_{i=a}^{b}(x_i - m)^s}{b-a} \tag{3}$$

Therefore, for a probability density function f(x), the first moment about the mean is always zero (with s = 1), while the second moment is the variance. The third central moment is defined as skewness such that a distribution skewed to the right has a positive value, while one shifted towards the left has a negative skewness.

The second most highly correlated feature is the mean peak distribution, which is defined as the mean distance between peaks for the logarithmic representation of the signal energy. The third feature is the number of non-zero values of the normalized log-energy signal.

Features 4-10, preceded by MFCC, are Mel-Frequency Cepstral Coefficients [20], which represent the spectral characteristics of the signal. A cepstrum is the result of the Inverse Fourier Transform of the logarithm of a signal spectrum. Mel-Frequency Cepstral Coefficients are based on the mel scale, which is a perpetual scale of pitches judged by listeners to be equidistant from one another [21]. The relationship between the frequency and mel scales is logarithmic, and can be defined by the following formula (though other variations exist) [21]:

$$MEL(f) = 2595 \cdot log_{10}(1 + \frac{f}{700}) \quad (4)$$

Assume we attempt to obtain the MFCC of a 0.25 second audio clip. The first step is to compute the Discrete Fourier Transform of the window, Y(k), from which we can obtain the power spectral density (PSD), P(k) using the following formula in which W is the window size:

$$P(k) = \frac{1}{W}|Y(k)|^2 \quad (5)$$

Next, we must obtain an estimate of the energies of different frequency ranges in the signal. However, the human ear can discern differences in frequency at low frequency ranges with a much higher resolution than at higher ranges, due to the physical properties of the cochlea. Therefore, a Mel Filterbank [22] is applied to the signal, which consists of N partially overlapping triangular window functions in the frequency domain. At higher frequencies, the triangular filters are wider, because we are less concerned with small variations in energy in these frequency ranges. [23]. Generally, N is a value between 20 and 40, with each window in the filterbank equally spaced in the Mel domain, which ranges from 300 Hz to 8000 Hz for speech-processing applications.

A dot product is computed between the filterbank and vector P(k), which yields N intermediary coefficients- one for each triangle window function in the filterbank. Because humans do not perceive loudness on a linear scale, the logarithm is calculated for all N coefficients. Finally, the Discrete Cosine Transform [24] (DCT) of the log powers is applied in order to decorrelate the energies of the overlapping filterbank energies. The resulting coefficients are used to extract statistical features as shown in Table III.

C. Smartwatch Feedback: A Survey

Figure 6 provides several pertinent questions from the survey. From the total sample of 221 respondents, 86% claimed to be right handed, 12% right-handed, and the remaining responded that they were 'unsure' or the question was 'not applicable'. In the following question, a total of 76% of respondents stated that they generally would wear a watch on their left hand, with an additional 19% who preferred to wear the watch on their right. The remaining 5% of those surveyed expressed no preference.

TABLE III. Partial List of Selected Features

Rank	Feature Name
1	pcm_LOGenergy_sma_skewness
2	pcm_LOGenergy_sma_meanPeakDist
3	pcm_LOGenergy_sma_nnz
4	mfcc_sma[0]_quartile3
5	mfcc_sma[0]_meanPeakDist
6	mfcc_sma[0]_nnz
7	mfcc_sma[1]_quartile2
8	mfcc_sma[1]_meanPeakDist
9	mfcc_sma[1]_peakMean
10	mfcc_sma[1]_amean

TABLE IV. Audio: Confusion Matrix (Random Forest)

	Predicted Outcome				
Swallow Type	Apple	Chips	Water	Talk	Recall
Apple	46	4	0	0	92%
Chips	6	44	0	0	88%
Water	1	0	49	0	98%
Talk	0	0	0	50	100%
Precision	86.7%	89.7%	100%	100%	

The next question asked respondents how they felt about wearing watches in general. Most individuals stated that they always wear a watch (38%). However, 23% claimed that they preferred not to wear a watch, 24% stated that they would not mind, and 14% stated that they like to wear a watch. Only 1% of individuals claimed that they would not consider wearing a watch. However, another survey question revealed that those who drank water out of a glass would use their primary hand to lift the cup for their mouth (69%), rather than the secondary hand on which the watch is worn (20%) with a remaining 10% claiming to be unsure. Clearly, this would pose a challenge to detection of liquid consumption.

The next question asked respondents if they would be willing to wear a watch on the opposite hand to which they are accustomed. The results were quite promising, with 40% of respondents answering 'maybe', 32% answering 'yes', and 28% answering 'no'. It appears that enough individuals are willing to change which hand they wear their watch, to make detection of most eating habits possible if the algorithm settings are customized to their personal habits.

VII. Conclusion and Future Work

This paper presents a novel approach to detecting ingestion of foods and liquids, using a Samsung smartwatch for identification of bites and swallows from acoustic signals. We conclude that the smartwatch platform is a strong choice for non-invasive evaluation of eating habits, and the versatility and comfort of the watch platform is a substantial advantage over existing schemes that rely on custom hardware solutions. This paper also presents a survey of users about smartwatch usage which confirms that a substantial portion of individuals would be willing to wear a watch on the hand with which they primarily eat.

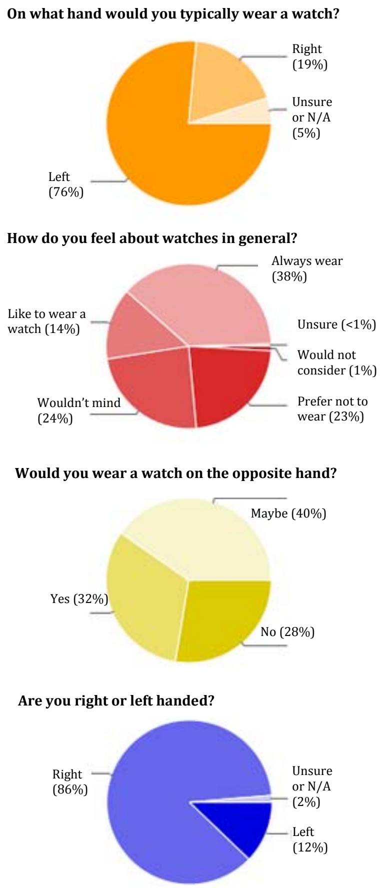


Fig. 6. Partial survey results are shown above.

In future works, we would like to automatically detect the hand on which the watch is being worn, and modify the classification thresholds accordingly in order to improve classification accuracy. This is necessary because the magnitude of the signal will vary if the watch is not worn on the same hand used to pick up an item of food. We would also like to explore the integration of audio-based detection of eating with inertial sensors for gesture recognition. Because smartwatches include an accelerometer and gyroscope, detection of eating-related motions coupled with audio data can improve our ability to characterize a meal.

ACKNOWLEDGMENT

This work is funded by the National Science Foundation AIR Option 1: Award Number 1312310

REFERENCES

[1] "Centers for disease control and prevention. annualmedical spending attributable to obesity: Payer-and service-specific estimates," http://www.cdc.gov/obesity/data/adult.html, accessed: 2015-02.

[2] P. S. Freedson, E. Melanson, and J. Sirard, "Calibration of the Computer Science and Applications, Inc. accelerometer," *Med Sci Sports Exerc*, vol. 30, no. 5, pp. 777–781, May 1998.

[3] P. S. Freedson, K. Lyden, S. Kozey-Keadle, and J. Staudenmayer, "Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample," *J. Appl. Physiol.*, vol. 111, no. 6, pp. 1804–1812, Dec 2011.

[4] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 1, p. 21, 2012.

[5] J. Fontana, J. Higgins, S. Schuckers, and E. Sazonov, "Energy intake estimation from counts of chews and swallows," *Journal of Appetite*, no. 85, pp. 14–21, 215.

[6] H. Kalantarian, N. Alshurafa, M. Pourhomayoun, S. Sarin, T. Le, and M. Sarrafzadeh, "Spectrogram-based audio classification of nutrition intake," in *IEEE EMBS Healthcare Innovations & Point of Care Technologies (HIPT)*, 2014.

[7] O. Amft, "A wearable earpad sensor for chewing monitoring," in *Sensors, 2010 IEEE*, Nov 2010, pp. 222–227.

[8] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, N. Sazonova, E. L. Melanson, and M. Neuman, "Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior," *Physiological Measurement*, vol. 29, no. 5, p. 525, 2008.

[9] M. Nagae and K. Suzuki, "A neck mounted interface for sensing the swallowing activity based on swallowing sound," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, Aug 2011, pp. 5224–5227.

[10] O. Amft, M. Kusserow, and G. Troster, "Bite weight prediction from acoustic recognition of chewing." *IEEE Trans. Biomed. Engineering*, vol. 56, no. 6, pp. 1663–1672, 2009.

[11] E. S. Sazonov, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. L. Melanson, and M. R. Neuman, "Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior," *IEEE Trans Biomed Eng*, vol. 57, no. 3, pp. 626–633, Mar 2010.

[12] S. Passler and W. Fischer, "Acoustical method for objective food intake monitoring using a wearable sensor system," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*, May 2011, pp. 266–269.

[13] W. Pler S and F. W. M, "Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food." *Physiological Measurement*, pp. 1073–1093, jun 2012.

[14] D. Brochetti, M. Penfield, and S. Burchfield, "Speech analysis techniques: A potential model for the study of mastication sounds," *Journal of Texture Studies*, vol. 23, no. 2, pp. 111–138, 1992. [Online]. Available: http://dx.doi.org/10.1111/j.1745-4603.1992.tb00515.x

[15] W. E. HI, A. E. Deibel, C. T. Glembin, and E. Munday, "Analysis of food crushing sounds during mastication: Frequency-time studies1," *Journal of Texture Studies*, vol. 19, no. 1, pp. 27–38, 1988. [Online]. Available: http://dx.doi.org/10.1111/j.1745-4603.1988.tb00922.x

[16] "Opensmile faq," http://www.audeering.com/research/opensmile, accessed: 2015-02.

[17] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: The munich versatile and fast open-source audio feature extractor," in *Proceedings of the International Conference on Multimedia*, ser. MM '10. New York, NY, USA: ACM, 2010, pp. 1459–1462. [Online]. Available: http://doi.acm.org/10.1145/1873951.1874246

[18] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[19] T. Pfister and P. Robinson, "Real-time recognition of affective states from nonverbal features of speech and its application for public speaking skill analysis," *Affective Computing, IEEE Transactions on*, vol. 2, no. 2, pp. 66–78, April 2011.

[20] S. Imai, "Cepstral analysis synthesis on the mel frequency scale," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'83.*, vol. 8. IEEE, 1983, pp. 93–96.

[21] D. O'Shaughnessy, *Speech communication: human and machine*. Addison-Wesley, 1987.

[22] B. Milner and X. Shao, "Speech reconstruction from mel-frequency cepstral coefficients using a source-filter model." in *INTERSPEECH*. Citeseer, 2002.

[23] "Mel frequency cepstral coefficient (mfcc) tutorial," http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/.

[24] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *Computers, IEEE Transactions on*, vol. 100, no. 1, pp. 90–93, 1974.

k-Nearest-Neighbour based Numerical Hand Posture Recognition using a Smart Textile Glove

Cai Li *, Katharina Bredies†, Anja Lund‡, Vincent Nierstrasz‡, Paul Hemeren* and Dan Högberg §
*Cognition and Interaction Lab, School of Informatics, University of Skövde, Sweden
Email: gauss.lee@his.se, paul.hemeren@his.se
†Department of Design, Faculty of Textiles, Engineering and Business, University of Borås, Sweden
Email: katharina.bredies@hb.se
‡Department of Textile Technology, Faculty of Textiles, Engineering and Business, University of Borås, Sweden
Email: anja.lund@hb.se, vincent.nierstrasz@hb.se
§School of Engineering Science, University of Skövde, Sweden
Email: dan.hogberg@his.se

***Abstract*—In this article, the authors present an interdisciplinary project that illustrates the potential and challenges in dealing with electronic textiles as sensing devices. An interactive system consisting of a knitted sensor glove and electronic circuit and a numeric hand posture recognition algorithm based on k-nearest-neighbour (kNN) is introduced. The design of the sensor glove itself is described, considering two sensitive fiber materials – piezoresistive and piezoelectric fibers – and the construction using an industrial knitting machine as well as the electronic setup is sketched out. Based on the characteristics of the textile sensors, a kNN technique based on a condensed dataset has been chosen to recognize hand postures indicating numbers from one to five from the sensor data. The authors describe two types of data condensation techniques (Reduced Nearest Neighbours and Fast Condensed Nearest Neighbours) in order to improve the data quality used by kNN, which are compared in terms of run time, condensation rate and recognition accuracy. Finally, the article gives an outlook on potential application scenarios for sensor gloves in pervasive computing.**

Keywords–smart textile controllers, kNN, piezoresistive and piezoelectric fibers

I. INTRODUCTION

Artificial Intelligence (AI) research into pervasive computing deals with intelligent systems that are usually highly distributed in space with a multitude of possible input and output modalities. In addition to static distributed sensors, mobile and wearable sensors have recently become increasingly important to gather more accurate and different information on the state and performance of the user [1], e.g., to track posture, movements, or even physical and emotional states. Due to the advances in textile technology and materials research, textile sensors have emerged as a new alternative to established electronic components in wearable applications. As Roggen et al. [2] point out, textile sensors are particularly useful in wearable applications for their close proximity to the skin, potential multimodality, convenience and wearing comfort, long-term use, and user acceptance.

However, textile sensors also pose a challenge on the computing part, as they tend to be much less predictable than conventional electronic components. In fact, textile sensing materials often produce a significant amount of noise; they are subject to a lot of mechanical stress and tend to wear out over time; their performance can be dependent on environmental influences such as temperature and humidity; and their physical structure is much more difficult to model and predict before the actual production, compared to standard sensors. The use of textile sensors in a pervasive computing system therefore relies on the use of an appropriate recognition algorithm that is able to process the sensor data in a meaningful way.

In this paper, we present an interactive system that uses a textile sensing device - that is, a glove with sensors on each finger - as a wearable controller in a pervasive computing context. Through the design of a complete e-textile system, we also would like to comprehend fundamental benchmarks for a sustainable and reusable textile sensing device. Gloves are a popular and well-explored form of a wearable textile that can be used for posture and gesture recognition to interact with computational systems [3]. Data gloves have been used for sign language recognition, robot control, graphic editor control, virtual environments, number recognition, television control, 3D modelling [4]. Our goal was to provide an integrated, low-level, low-cost alternative to more accurate and sophisticated sensor gloves. Unlike most commercial systems, this system uses sensors from textile materials that are fully integrated with the surrounding structure, and therefore both lightweight and comfortable as well as cheap and easy to produce. As such, they provide a robust and mobile solution which is less sensitive to distance and lighting conditions than camera-based recognition systems, adaptable, and less complicated than commercial system with a higher sensor density and accuracy. We also explain the adaptation of existing gesture recognition algorithms for the use with the new textile sensors. We conclude with a test case to evaluate the combination of textile hardware, wearable setup and a recognition algorithm.

In section II-A, we give a brief overview of the hardware components of the interactive system we used, from hardware (sensor glove) to algorithm. We then introduce an example application for numeric hand posture recognition, for which we deployed a k-nearest-neighbour (kNN) approach, in section II-B. The experiments to compare the two kNN models based on both simulation and sensory data are described in section III together with a performance demonstration of the algorithm with a NAO robot. In Section IV, we summarize the perspective of smart textile technology used for activity recognition. We also conclude with an assessment of the challenges of applying e-textile sensors in our sensor glove application.

 ISBN: 978-1-61208-421-3

II. The Glove Control System

A. Hardware Setup

Data gloves are a popular application for wearable sensors, as they provide wearer comfort while enabling the monitoring of hand movements and postures. Existing projects have used a broad range of sensor technologies to detect finger bending, such as commercial bending sensors [5], [6], [7], optical fibers [8], [9], or printed polymer sensors [10]. The textile, in these cases, merely serves as a carrier substrate for the sensor components. In contrast, our work demonstrates bend sensitive gloves where textile fibres with sensor properties were directly integrated in the textile during the knitting process (Figure 1), resulting in a comfortable and lightweight construction. Unlike existing systems, these gloves are fabricated as one piece, with the sensors being part of the knitted structure. This approach is more similar to [11], where the sensor yarn is integrated into a hand-crocheted glove. Being fully automated, the production process and the appearance of the glove corresponds to a normal knitted glove, resulting in a lightweight and cheap wearable that could be produced on a large scale. For our glove, we considered two e-textile sensor materials, piezoresistive and piezoelectric fibers.

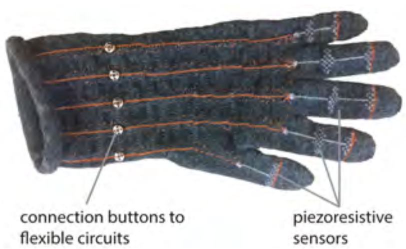


Figure 1. The e-textile glove used in our work. The light grey fibres are piezoresistive fibres. The buttons are sewn on the textiles to connect to the flexible circuit board.

1) Piezo-resistive material: Piezo-resistive materials change their resistance when deformed, i.e., pressed or stretched. While this is essentially true for all conductive materials, the material's selection for a certain application will also depend on the material's mechanical properties and its suitability for the relevant manufacturing process. Textile piezo-resistive sensors may be constituted by, e.g., silicon-based coatings filled with carbon particles [12], carbon-coated rubbery fibres [13] or conductive fibers arranged in a stretchable and elastic textile structure [14]. In our project, we use a piezo-resistive thread (Bekaert Bekinox©50/2 [15]) that is a blend of 20% short steel fibers and 80% polyester yarn with an average conductivity of $50\Omega/\text{cm}$ under strain that has been used in a similar project to produce bend sensors on a glove [11]. When in a relaxed state, only a few steel fibers make contact within the thread, resulting in a high resistance. When the thread is stretched, the steel fibers are forced closer to each other, which increases the conductivity of the material. When used as sensors, the resistance of a piezoresistive fiber is continuously measured and will be proportional to the amount of pressure or strain.

This piezoresistive thread is readily available, relatively cheap, easy to work with, making it simple to use as a variable resistor in a simple voltage divider circuit, where the voltage drop over it is proportional to its resistance. The material is not insulated and surface contact is sufficient for electrical connections. The piezoresistive effect, however, depends highly on the production conditions and composition of the sensors in a particular object: It can be influenced by the yarn tension on the knitting machine, by the density of the knitted structure, the stretchability and elasticity of the surrounding material, as well as the fit of the wearable it is part of. Also, the resistance of the material is considerably high, adding up to a resistance of up to $1\text{M}\Omega$ for a single sensor in relaxed state.

2) Piezo-electric fibers: Piezoelectric materials generate an electric voltage when deformed. This property is present in different types of materials, e.g., minerals, ceramics and polymers, and is due to a persistent polarisation in the molecular structure, which causes a change in the charge density in response to deformation. This can be measured as a transient voltage across the material's boundaries. The continuous production of the piezoelectric polymer fibers used in this work was recently presented [16]. This fibre (produced and kindly supplied by Swerea IVF, Mölndal, Sweden) can be readily processed in standard textile production methods, e.g., knitting, weaving, embroidery, and is highly sensitive to strain. For example, a textile band woven from this fibre and fastened around the chest of a person, has been shown to generate clear output signals in response to the wearer's heartbeat [17].
An advantage of piezoelectric materials is that, as opposed to the piezo-resistive ones, they generate their own voltage. In practice though, especially in polymers, the generated current is extremely small and the piezoelectric fiber must be connected to an operational amplifier working as a high impedance buffer, for the output signals to be of a useful amplitude. Thus, in comparison to the piezoresistive fiber, the electronic assembly for the piezoelectric material is more complex and requires more components.

3) Robot and its software interface: The humanoid used in our work is a commercialised robot from Aldebaran Robotics, called NAO [18]. The NAO robot has 25 degrees of freedom and multiple useful sensors (e.g., ultrasound, gyro and motor sensors). It can also perform a lot of sophisticated functions, such as dancing, walking and speaking. The software embedded in the robot is called Naoqi which works as a mid-ware to synchronise all the modules running on the NAO. In our work, the hand posture recognition module and the entertainment module are being synchronised for communicating with each other based on hand postures.

B. Data Processing

The piezoresistive sensors in the glove show a complex response to bending of one or several fingers at a time that does not correspond with a straightforward proportional increase in conductivity. This behaviour is somewhat typical for wearables. Most applications of wearable devices involve a sophisticated process of translating sensory data into context specific meanings based on a variety of computational models [2]. In order to interpret data from wearable sensors, many different pattern recognition and machine learning techniques can be used, such as neural networks [19], fuzzy logic models [20], dynamic time warping [21] and knowledge based models [22]. Generally speaking, these techniques are usually employed in an activity recognition chain which includes

 ISBN: 978-1-61208-421-3

sequential functions of data preprocessing/segmentation, feature extraction, classification, classifier fusion, decision filtering and high-level reasoning[2]. The activity recognition chain is a salient part of the whole e-textile system consisting of a three-level design: sensor hardware, signal processing/activity analysis and high-level interaction[2]. The sensor hardware design refers to the process of sensor design and characteristics test. After a proper sensor test and signal preprocessing (which guarantees that the sensor is appropriately designed and signals are not noisy), acquired data can then be used in an activity recognition chain (ARC). In the end of an ARC, a high level interactive model (if required) might be designed for more sophisticated applications (e.g., emotion recognition, cognitive processes).

C. kNN: an algorithm for hand posture recognition

Generally speaking, there are three categories of training/learning algorithms broadly used for recognizing different hand postures with high accuracy: neural networks (NN), hidden markov models (HMM) and instance-based learning models (kNN) [3]. However, there is no guarantee that NN or HMM can converge if structure configurations of NN or HMM are inappropriately set (e.g., number of layers for NN, number of hidden states for HMM) [3]. kNN can avoid no-convergence risk as there is always a classification decision based on calculation of k nearest neighbours. We therefore choose an instance based model (kNN) because of its simplicity for implementation. Also, as an unsupervised learning technique [23], kNN is data driven, which means it has the ability of continuous training with more data. This provides an easy approach to calibrating for new users by involving their data in a continuous training process.

1) kNN algorithms: k nearest neighbours is an algorithm that classifies a new dataset x based on a training data D. x can be a dataset with m dimensions and D is a labeled database (all the datasets have been correctly labeled with classes) containing n datasets. A normal kNN should include two general steps for classifying an input dataset [24]: *a) calculate the k nearest datasets within D for the input dataset x. b) return the class that represents the maximum of the k datasets.* The nearest neighbours can be determined by the calculation based on distinct distance metrics, such as euclidean distance, minkowski distance and mahalanobis distance [24]. In our work, we use the simplest euclidean distance.

Since the distance calculation dependent on training database D directly determines the class of input dataset x, the quality of the database becomes a salient factor for kNN based classifiers. There are two factors that might potentially affect the quality of the database: a) training data size (n) and b) dimensions or attributes of data (m). Obviously, if the training data size is too large, it will deteriorate the speed of distance calculation in real time, causing the failure of algorithm implementation. In order to avoid this, a data condensation technique is needed to remove the redundancy in the training data, which then leaves the minimum number of data for sketching the probabilistic distribution of the original data [25]. On the other hand, kNN still suffers the curse of dimensionalities [26]. A dataset with too many dimensions or attributes can cause the failure of kNN classification. Solutions for this problem are using dimension reduction techniques, e.g., principal component analysis [23] and backward elimination [26]. Meanwhile, in order to have usable training data, some preprocessing techniques are necessary for standardizing the data, e.g., signal filtering (remove noise in the data), signal segmentation and normalization [3].

2) Data preprocessing: In our work, data preprocessing only involves filtering and normalization. The aim of filtering is to maximize the signal to noise ratio so that the influence of noise can diminished. The filtering has been fulfilled in electronic circuits by using standard low-pass resistor capacitor (RC) filters. Normalization is calculated on each dataset following $x'_D = \frac{x_D - min(x_D)}{max(x_D) - min(x_D)}$, where x_D and x'_D are a dataset before and after normalization, respectively. $min(x_D)$ is the minimum sensor value 0 and $max(x_D)$ is determined by the sensor value with users fully bending each finger. Then datasets are collected for postures corresponding to each respective number and captured variation when pivoting a user's wrist.

3) Data condensation: Since each dataset used in our work only contains five necessary numbers/dimensions from five fingers, the data dimension can be considered to be minimum so that dimension reduction algorithms are not used in our work. However, the size of the training data used in our work is 2000 for each posture which contains 1500 and 500 datasets for training and testing respectively. It is too large for real time implementation. Therefore, data condensation is necessary. The aim of data condensation is to find a subset of the original training data, which does not influence classification results [25]. There are a lot of types of data condensation algorithms. According to a complete survey of different data condensation techniques [27], CNN and FCNN outperform most of other algorithms with the good features of condensation rate and computation complexity. Therefore, we apply both algorithms on simulated gaussian data and hand posture data for comparison(for algorithm details, please refer to [28] and [29]).

III. Experiments and Results

In this section, we apply kNN algorithms on the sensor glove for numeric hand posture recognition. In order to clearly compare the two data condensation techniques used in the experiments, we first test their performance on a database containing three gaussian generated data classes. Then, we apply the two algorithms to the numeric hand posture database. Finally, a human-robot interaction demonstration is set up to show that a humanoid can be controlled by recognising numeric hand postures as "menu selection" input.

A. The glove textile and circuit construction

To knit a glove with each material, a Shima Seiki SWG091N industrial knitting machine with six yarn feeders has been used. The machine and the accompanying programming software provide templates for gloves that can be customized for specific hand measurements. The knitting process of a five-finger glove first produces the index, middle finger, ring finger and little finger separately, then proceeds with the upper hand part, adds the thumb and finishes the glove. This sequence requires that after finishing a finger, all yarns are automatically cut. The amount of yarn feeders limits the amount of yarns that can be used at the same time in one knit structure.

 ISBN: 978-1-61208-421-3

Given these constraints, the design of the sensors on the glove is limited to the length of the fingers and cannot easily be extended over the back of the hand. For the piezoelectric fiber, it would require re-connecting the fiber cores between the finger and the back of the hand; for both fiber types, it would be necessary to have more than six yarn feeders - one for each finger including the thumb, one for the basic material (wool), and one for the high-conductive copper thread (7/1 high flex copper thread from Karl Grimm [30]) that is used for the connections to each finger on the palm side of the hand.

Both sensor types are to a certain degree ambiguous in their response - they do not only react to stretching (i.e., bending of the fingers), but also to pressure (e.g., pressing the knuckles without bending the fingers). The two fiber types provide different kinds of signals: The piezoresistive fiber allows for continous readings of the amount of bending. The piezoelectric fiber produces an event-based signal depending on the force and velocity of the bending motion. For the posture recognition, we decided to work with a glove version that had piezoresistive sensors on all five fingers (Figure 1). While sensors and - to some extent - the electric connections can be very well integrated with the knitted structure of the glove, the more delicate electronic components of a wearable device usually have to be arranged on a more conventional substrate and then attached to the e-textile device through mechanical connections. In our case, all components other than the sensors and the connection to the sensor were mounted on a custom made printed flexible circuit board (C.I.F. AN10 1-sided plain flex circuit board with 35 microns copper layer), which was attached to the glove with snap buttons and closed around the wrist like a bracelet (Figure 2). This construction principle, which has been developed by [31], makes the sensing and communication circuit small and lightweight so that it can be conveniently used.

For both sensor gloves, we used an ATTiny 84 micro-controller as processor to read the sensor data from either the voltage divider or the amplifier (LMC 660 op amp) with a simple program that outputs the sensor data to a serial connection. The connection runs via a six-pin header that can be connected to a Bluetooth device or a FTDI board with a mini USB plug with the same footprint. The circuit can run on the power from the FTDI or on a LiPo coin cell battery placed in a small pouch on the flex circuit itself. For the piezoresistive glove, large capacitors (1μF) have been added parallel to the pullup resistors for filtering. 1).

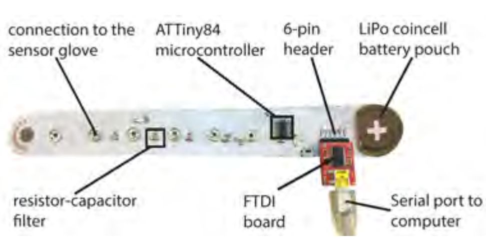


Figure 2. The flexible circuit wearable on one's wrist with component names and labels.

B. Test on simulation data

In order to visualise how the two data condensation techniques perform, we chose to test them with a simulation training database with three gaussian generated classes. Each data set contained a two dimensional coordinate in a $x - y$ space and each class contained 500 datasets. The testing database included 300 data of which each class had 100 datasets. The condensed database based on CNN and FCNN is shown in Figure 3.

From results shown in Figure 3, we clearly see that the FCNN

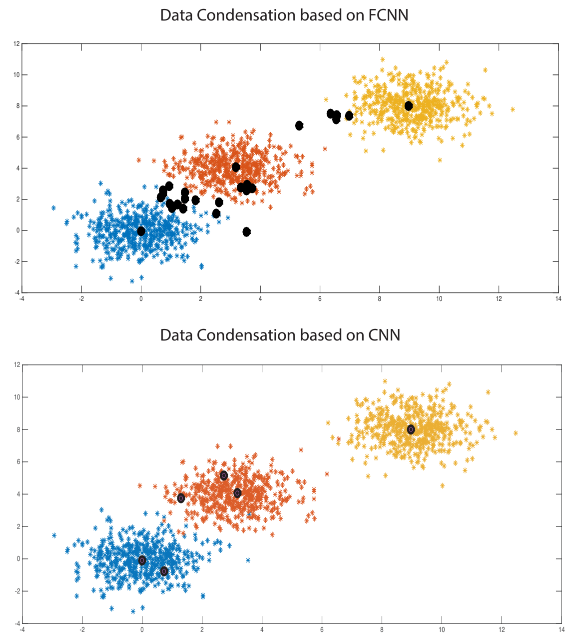


Figure 3. Upper: original database (shown in blue, red and yellow for three classes) and FCNN condensed database (black). Down: original database (shown in blue, red and yellow for three classes) and CNN condensed database (black).

condensed database is more aligned close to the borderlines separating every two classes except for the centroids. However, the CNN condensed database contains less data and is more distributed within the class clusters. In machine learning, finding class borderlines is very useful and important for classifiers such as a support vector machine, regression models, neural network and kNN[23]. Also it is more convenient to classify input data based on an FCNN condensed database with clear class boundaries. When verifying the trained model with the kNN rule (k = 3) based on the testing database, the accuracy rate is 98% and 93% for FCNN and CNN condensed data respectively. Clearly, in the case of simulation data, FCNN outperforms CNN in terms of accuracy. However, it takes less than 2 minutes to run CNN, compared to 10 minutes for FCNN.

C. Hand posture recognition experiments

The numeric hand postures (from 1 to 5) adopted in our experiments are shown in Figure 4. Using the sensor glove, 8000 datasets was collected from one user; each posture corresponds to 2000 datasets. For each posture, the database was split into 1500 for training and 500 for post-training test. Each data

 ISBN: 978-1-61208-421-3

contained five dimension input from each finger respectively. Then, CNN and FCNN were used to extract condensed data out of 6000 datasets. We assume that a calibration process (retraining) needs to be done for each user since every user has different hand size. The results of the comparison between CNN and FCNN are shown in Table I.

Figure 4. Hand postures representing numbers from 1 to 5 (from left to right).

TABLE I. Comparison of CNN and FCNN

Algorithms	Training Time (seconds)	Condensed rate	Accuracy
CNN	127.543s	94.43%	75.4 %
FCNN	5 hours (approximately)	87.5%	96.3 %

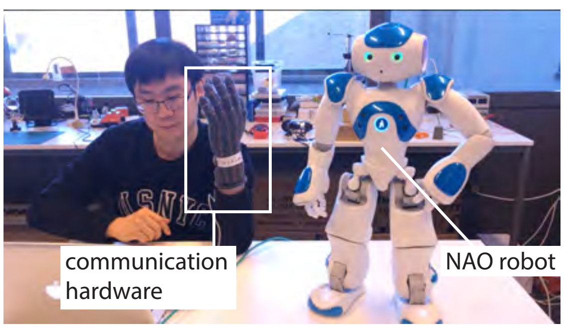


Figure 5. The sensor glove used in a human robot interaction demonstration.

From the results in Table I, we can draw some conclusions about the comparison of the two condensation techniques: a) in terms of condensation rate, CNN tends to remove more data than FCNN while FCNN can classify with much higher accuracy. There might be a trade off between condensation rate and classification accuracy. b) The training time of CNN is rather shorter than FCNN as FCNN is an incremental algorithm which involves a thorough database search for every incremental step. This result is similar to the results obtained by Amal et al [27] pertaining to the advantages and disadvantages of different data condensation techniques. With condensed datasets, kNN algorithms can run very fast. The response time for a new posture changed from a previous one is on average smaller than 0.01s compared to 1.2s for non-condensed datasets.

D. A robotics demonstration

Finally, the use of a self-designed sensor glove as an external controller to interact with a humanoid has been demonstrated. This human robot interaction relied on a previously trained classifier in real time, requiring the classifier to be both accurate and fast-responding. Three games were designed for the user to try on the NAO robot (Figure 5). The user could use hand gestures to communicate with the robot both to select which game to play, and to actually play it. For example, in the game named "number reaction", the user needed to react with hand postures to a number said by the NAO robot. This fully demonstrates the function of kNN model working with the sensor glove. As a result, the well-trained classifier could quickly determine hand postures in real time from data streams (Please refer to the video [32]).

IV. Conclusion

A. Challenges for our e-textile glove system

In this article, we demonstrated a smart textile glove used as a controller for an application in a robotics control context. Presumably, the smart textile glove can also be applied to control other pervasive devices which are communicable and controllable to our systems. However, there are still some challenges in our system, due to the variability in the characteristics of textile sensors:

- Sensitivity – In this particular e-textile system, the upper finger knuckles were not covered by the sensors, which reduces the sensitivity of the sensors for detecting finger bending. The extent and placement of the sensors can be improved by elongating the sensor to cover all finger knuckles and to use a looser fit for the glove, e.g., by adding elasthane to the non-conductive basis material.
- Hysteresis– Textile materials typically exhibit stress relaxation and creep, causing the sensitivity to degrade after a certain period of constant use. This fact limits the textile materials useful in sensor applications to those with high elasticity and good ageing properties, but also inevitably introduces a factor of time dependence in all measurements. One solution might be setting up a periodic test for the sensor glove to measure and record the variation of the sensors, and establishing a model to statistically describe the variation in order to compensate/cancel effects of deterioration.
- Offset – A textile sensor will typically have a pre-strain, depending among other things on the size of the wearer. Integrating data from a large number of users is of great importance to establish a general calibration system making the sensor glove easily reusable. In our system, we use a retraining strategy to solve this problem. Obviously, retraining for every user will increase calibration time. A solution to this problem might be to statistically model the probabilistic variation of data from different users and establish a general calibration system which has the ability to integrate new data to reduce calibration time.

B. Future work

In future work, the whole sensor glove system needs to be improved from three different perspectives: a) Improvement on hardware design. Considering the constraints of the knitting machine, different knitting patterns and techniques must be tried to provide a more meaningful sensor allocation on the back of the hand. This might be achieved by combining piezoelectric fibers (to detect the bending motion) with piezoresistive fibers (to detect the posture). b) Improvement on the activity recognition chain. Calibrating the "ground truth" of the sensor glove is necessary to establish an accurate model for different hand gestures. This needs a detailed model of hand kinematics by clarifying the variation of sensor sensitivity and users' different hand sizes. The final aim of ground truth modelling is to accurately map one user's hand motion into a 3-dimensional Cartesian space. c) Improvement on applications. We assume that a simple numeric hand posture recognition system can be extensively developed to a sign language interpretation system. This is useful for helping speechless people to communicate with an intelligent machine. This application upgrade involves not only improvement on hardware by increasing the sensor functionality, but also by improving techniques for interpreting dynamic hand gestures in a sequence instead of only recognizing static hand postures.

ACKNOWLEDGEMENTS

The financial support of Västra Götalandsregionen (VGR), grant number RUN 612-0197-13, is gratefully acknowledged.

REFERENCES

[1] E. Sazonov and M. R. Neuman, Eds., Wearable Sensors. Academic Press, 2014.

[2] D. Roggen, G. Tröster, and A. Bulling, "Signal processing technologies for activity-aware smart textiles," in Multidisciplinary Know-How for Smart-Textiles Developers, T. B. T. M. K.-H. f. S.-T. D. Kirstein, Ed. Woodhead Publishing, 2013, ch. 12, pp. 329–365.

[3] J. J. LaViola Jr., "A Survey of Hand Posture and Gesture Recognition Techniques and Technology," NSF science and Technology Center, Tech. Rep., Jun. 1999.

[4] R. Zaman Khan, "Hand Gesture Recognition: A Literature Review," International Journal of Artificial Intelligence & Applications, vol. 3, no. 4, 2012, pp. 161–174.

[5] L. C. Pereira, R. V. Aroca, and R. R. Dantas, "FlexDGlove: A Low Cost Dataglove with Virtual Hand Simulator for Virtual Reality Applications," in Proceedings of the 2013 XV Symposium on Virtual and Augmented Reality, ser. SVR '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 260–263.

[6] S. Kang, J. Kim, and J. Lee, "Intuitive Control Using a Mediated Interface Module," in Proceedings of the International Conference on Advances in Computer Enterntainment Technology, ser. ACE '09. New York, NY, USA: ACM, 2009, pp. 435–436.

[7] H. Perner Wilson, "The Gloves Project," 2015, [Retrieved: 11-06-2015]. [Online]. Available: http://theglovesproject.com

[8] R. C. King, L. Atallah, A. Darzi, and G. Z. Yang, "An HMM Framework for Optimal Sensor Selection with Applications to BSN Sensor Glove Design," in Proceedings of the 4th Workshop on Embedded Networked Sensors, ser. EmNets '07. New York, NY, USA: ACM, 2007, pp. 58–62.

[9] E. Fujiwara, D. Y. Miyatake, M. F. M. Santos, and C. K. Suzuki, "Development of a Glove-based Optical Fiber Sensor for Applications in Human-robot Interaction," in Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction, ser. HRI '13. Piscataway, NJ, USA: IEEE Press, 2013, pp. 123–124.

[10] N. Tongrod, A. Tuantranont, and T. Kerdcharoen, "Design and Development of Data-glove Based on Printed Polymeric Sensors and Zigbee Networks for Human Computer Interface," in Proceedings of the 4th International Convention on Rehabilitation Engineering & Assistive Technology, ser. iCREATe '10. Kaki Bukit TechPark II,, Singapore: Singapore Therapeutic, Assistive & Rehabilitative Technologies (START) Centre, 2010, pp. 47:1—-47:4.

[11] M. Satomi, L. Worbin, and B. Scholtz, "Textile Resistance," Borås, Sweden, 2011, [Retrieved: 11-06-2015]. [Online]. Available: http://hdl.handle.net/2320/10180

[12] L. Guo, L. Berglin, U. Wiklund, and H. Mattila, "Design of a garment-based sensing system for breathing monitoring," Textile Research Journal, vol. 83, no. 5, 03 2013, pp. 499–509.

[13] F. Lorussi, W. Rocchia, E. Scilingo, A. Tognetti, and D. De Rossi, "Wearable, redundant fabric-based sensor arrays for reconstruction of body segment posture," Sensors Journal, IEEE, vol. 4, no. 6, Dec 2004, pp. 807–818.

[14] L. Guo, L. Berglin, and H. Mattila, "Improvement of electro-mechanical properties of strain sensors made of elastic-conductive hybrid yarns," Textile Research Journal, vol. 82, no. 19, 11 2012, pp. 1937–1947.

[15] Bekaert, "Bekaert Bekinox Stainless Steel Fibers for Shielding Textiles," 2014, [Retrieved: 11-06-2015]. [Online]. Available: http://www.bekaert.com/en/products/basic-materials/textile/stainless-steel-fibers-for-anti static-textiles-bekinox

[16] B. Hagström, A. Lund, and E. Nilsson, "Method of producing a piezo-electric and pyroelectric fiber," 2014, [Retrieved: 11-06-2015]. [Online]. Available: http://www.google.com/patents/WO2014161920A1?cl=en

[17] E. Nilsson, A. Lund, C. Jonasson, C. Johansson, and B. Hagström, "Poling and characterization of piezoelectric polymer fibers for use in textile sensors," Sensors and Actuators A: Physical, vol. 201, no. 0, 2013, pp. 477 – 486.

[18] D. Gouaillier and P. Blazevic, "A Mechatronic Platform, The Aldebaran Robotics Humanoid Robot," in IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics. IEEE, Nov. 2006, pp. 4049–4053.

[19] G. B. Ramaiah, R. Y. Chennaiah, and G. K. Satyanarayanarao, "Investigation and modeling on protective textiles using artificial neural networks for defense applications," Materials Science and Engineering: B, vol. 168, no. 1-3, Apr. 2010, pp. 100–105.

[20] X. Luo, W. Hou, Y. Li, and Z. Wang, "A fuzzy neural network model for predicting clothing thermal comfort," Computers & Mathematics with Applications, vol. 53, no. 12, Jun. 2007, pp. 1840–1846.

[21] P. Tormene, T. Giorgino, S. Quaglini, and M. Stefanelli, "Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation." Artificial intelligence in medicine, vol. 45, no. 1, Jan. 2009, pp. 11–34.

[22] L. E. Dunne, P. Walsh, S. Hermann, B. Smyth, and B. Caulfield, "Wearable Monitoring of Seated Spinal Posture," IEEE Transactions on Biomedical Circuits and Systems, vol. 2, no. 2, Jun. 2008, pp. 97–105.

[23] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[24] L. Kozma, "k Nearest Neighbors algorithm (kNN)," Helsinki University of Technology, Tech. Rep., 2008, [Retrieved: 11-06-2015]. [Online]. Available: http://www.lkozma.net/knn2.pdf

[25] G. Gates, "The reduced nearest neighbor rule (Corresp.)," IEEE Transactions on Information Theory, vol. 18, no. 3, May 1972, pp. 431–433.

[26] S. Deokar, "Weighted k nearest neighbor," University of Minnesota Twin Cities, Tech. Rep., 2009.

[27] B.-A. A. R. Miloud-Aouidate Amal, "Survey of Nearest Neighbor Condensing Techniques," International Journal of Advanced Computer Science and Applications(IJACSA), vol. 2, no. 11, 2011, pp. 59–64.

[28] O. Sutton, "Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction," 2012, pp. 1–10.

[29] F. Angiulli, "Fast nearest neighbor condensation for large data sets classification," IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. November, 2007, pp. 1450–1464.

[30] K. Grimm, "Karl Grimm Fabrik leonischer Waren," 2015, [Retrieved: 11-06-2015]. [Online]. Available: http://www.karl-grimm.com

[31] H. Perner Wilson, "Gauntlet Circuit Design," 2015, [Retrieved: 11-06-2015]. [Online]. Available: http://www.plusea.at/?p=5168

[32] C. Li, "Demo for a Smart Textile Glove in human robot interaction ," 2015, [Retrieved: 11-06-2015]. [Online]. Available: https://www.youtube.com/watch?v=-aKF_EQHEbk

Component Templates and Service Applications Specifications to Control Dynamic Adaptive System Configurations

Holger Klus

ROSEN Technology & Research Center GmbH
Lingen (Ems), Germany
email: hklus@rosen-group.com

Andreas Rausch, Dirk Herrling
Department of Informatics
Technische Universität Clausthal
Clausthal-Zellerfeld, Germany
email: {andreas.rausch, dirk.herrling}@tu-clausthal.de

***Abstract*—Dynamic adaptive systems are systems that change their behavior at run time, based on system, user, environment and context information and needs. System configuration in terms of structure and behavior of open, self-organized systems cannot completely be predicted beforehand: New components may join, others may leave the system, or the behavior of individual components of the system may change over time. But in many cases, it is necessary to ensure the compliance of the resulting overall system configuration to users, environment and context requirements. Therefore we have elaborated an approach to specify those requirements based on so called component templates and service application specifications. These specifications can be described without the necessity to know individual components, their specific interfaces or possible system configurations. Thus, we can control the resulting system configurations of an open, self-organizing system with respect to users, environment and context requirements. Our approach has been implemented on top of our component model and corresponding platform implementation called Dynamic Adaptive System Infrastructure (DAiSI).**

Keywords—self-adaptation; dynamic adaptive systems; component templates and service application specification; adaptive component model; decentralized configuration.

I. Introduction

Software-based systems pervade our daily life—at work as well as at home. Public administration or enterprise organizations can scarcely be managed without software-based systems. We come across devices executing software in nearly every household. The continuous increase in size and functionality of software systems has made some of them among the most complex man-made systems ever devised [1].

In the last two decades, the trend towards “everything, every time, everywhere” has been dramatically increased through a) smaller mobile devices with higher computation and communication capabilities, b) ubiquitous availability of the Internet (almost all devices are connected with the Internet and thereby connected with each other), and c) devices equipped with more and more connected, intelligent and sophisticated sensors and actuators.

Nowadays, these devices are increasingly used within an organically grown, heterogeneous, and dynamic IT environment. Users expect them not only to provide their primary services but also to collaborate autonomously with each other and thus to provide real added additional value. The challenge is therefore to provide software systems that are correct, stable and robust in the presence of increasing challenges such as change and complexity [5].

Change is inherent, both in the changing needs of users and in the changes, which take place in the operational environment of the system. Hence, it is essential that our systems are able to adapt to maintain the satisfaction of the user expectations and environmental changes in terms of an evolutionary change [2].

Dynamic change, in contrast to evolutionary change, occurs while the system is operational. Dynamic change requires that the system adapts at run time. Therefore we must plan for automated management of adaptation. The systems themselves must be capable of determining what system change is required and initiate and manage the change process wherever needed. This is the aim of self-managed systems [3].

Self-managed systems are those capable of adapting to the current context as required through self-configuration, self-healing, self-monitoring, self-tuning, and so on. These are also referred to as self-x, autonomic systems. Additionally, new components may enter or leave the system at run time. We call those systems ‘dynamic adaptive’ systems [4].

Providing dynamic adaptive systems is a great challenge in software engineering [5]. In order to provide dynamic adaptive systems, the activities of classical development approaches have to be partially or completely moved from development time to run time. For instance, devices and software components can be attached to a dynamic adaptive system at any time. Consequently, devices and software components can be removed from the dynamic adaptive system or they can fail as the result of a defect. Hence, for dynamic adaptive systems, system integration takes place during run time.

To support the development of dynamic adaptive systems a couple of infrastructures and frameworks have been developed, as discussed in a related work section, Section II. In our research group, we have also developed a framework for dynamic adaptive (and distributed) systems, called DAiSI. DAiSI is a service-oriented and component based platform to implement dynamic adaptive systems [6].

Based on the existing components and their provided and required services DAiSI is able to autonomously find and es-

 ISBN: 978-1-61208-421-3

tablish during run time valid system configurations with respect to specific optimization goals and system guarantees. Even if new components join, or others leave DAiSI, or the behavior of individual components within DAiSI changes over time, DAiSI is able to reconfigure the overall system and establish a new valid system configuration at run time.

But in many cases, it is necessary to ensure the compliance of the resulting overall system configuration to users, environment and context requirements. Therefore we have elaborated an approach to specify those requirements based on so called component templates and service application specifications. The basic idea of our approach is to specify services users or the environment may be interested in form of so called "service application" specifications.

A service application specification consists of a set of so called "component templates". Each template is a placeholder for a set of components with specific properties. The application developer specifies the properties of component templates and service applications during design time. During run time, DAiSI tries to establish the required service applications by assigning autonomously existing components to compatible templates in the corresponding service application specifications.

These specifications can be described without the necessity to know individual components, their specific interfaces or possible system configurations. We have successfully implemented the component templates and service applications specifications on top of the existing component model of DAiSI. Thus, DAiSI is not only able to find and establish valid system configurations but also to find and establish them with respect to users, the environment and context requirements, which have to be explicitly expressed in component templates and service applications specifications.

The development of DAiSI was always motivated through running application examples and demonstrators. As DAiSI has been developed for more than ten years, we have demonstrated the application of our approach and our infrastructure in a couple of different research demonstrators and industrial prototypes and products.

The rest of the paper is structured as follows: In Section II, we present other works, we see as related to the DAiSI and its newest additions. Section III presents the fundamentals of DAiSI as it was prior to the additions, presented in this paper. Section IV describes a small sample application we use to illustrate the need to control possible system configurations in dynamic adaptive systems. Section V presents an approach to describe valid system configurations with regard to applications. In Section VI, we provide a notation for the specification of application requirements. Section VII presents an algorithm that leads to a requirements conform system configuration and explains why DAiSI only produces valid system configurations during run time with respect to users, environment and context requirements. A short conclusion will round the paper up.

II. Related Work

Component-based software development, component models and component frameworks provide a solid approach to support evolutionary changes to systems. It is a well-understood method that proved useful in numerous applications. Components are the units of deployment and integration. This allows high flexibility and easy maintenance. During design time components may be added or removed from a system [7].

However, the early component models did not provide means of adding or removing components from a running system. Also, the integration of new interaction links (e.g., component bindings) was not possible. Service-oriented approaches stepped up to the challenge. These systems usually maintain a service repository, in which every component that enters the system is registered. A component that wants to use such a component can query the service register for a matching service and connect to it, if one is found. For the domain of dynamic systems this means that a component can register its provided and required services. If a suitable service provider for one of the required services registers itself, it can be bound to satisfy the required service [8].

Service-oriented approaches have the inconvenient characteristic of not dealing with the adaptability of components. A component developer is solely responsible for the implementation of the adaptive behavior. This starts at the application logic and stretches to the discovery of unresponsive services, the discovery of a newly available service, the discovery of services with a better quality of service, and so on. A couple of frameworks have been developed to support dynamic adaptive behavior, while, at the same time, making it easier for the developer to focus on implementing the behavioral changes in his component.

REX is a framework for the support of dynamic-adaptive systems. It used the experience gained in the research for CONIC [9] and aimed at dynamic adaptive, parallel, distributed systems. The concept was that such systems consist of components that are linked by interfaces. A new interface description language was invented, to be able to describe the interfaces. Components were seen as types, allowing multiple instances of every component to be present at run-time. Just like CONIC, REX allowed the creation and termination of component instances and the links between them. Both, CONIC and REX share the disadvantage that they support dynamic reconfiguration only through explicit reconfiguration programs. These need to be different for every situation that is detected and intended. The approach moves the adaptation logic out of the component, but nevertheless, the developer has to deal with the adaptation strategy for every possible occurring change [10][11].

Current frameworks such as ProAdapt [12] and Config.NETServices [13] have a more generic adaption and configuration mechanism. Components that were not known during the design-time of the system can be added or removed from the dynamic adaptive system during run-time. Therefore, the framework provides a generic component configuration mechanism. As with our first version of the DAiSI framework, these frameworks are based on a centralized configuration mechanism. Moreover, the underlying component model is restricted—for instance the exclusive usage of services cannot be described.

In [14], the authors presented a solution to ensure syntactical and semantical compatibility of web services. They used the Web Service Definition Language (WSDL) and enriched it with the Web Service Semantic Profile (WSSP) for the semantical information. Additionally they allowed an application architect to further reduce the configuration space through the specification of constraints. While their approach is able to solve the sketched problem of preventing the wiring of components that should not be connected, they only focus on the service definition and compatibility. Our DAiSI approach defines an infrastructure in which components are executed that implement a specific component model. We do want to compose an application out of components that can adapt its behavior at run-time. We achieve this by mapping sets of required services to sets of provided services and thus specifying which provided services depend on which required services. The solution presented in [14] does not offer a component model. All rules regarding the relation between required and provided services would have to be specified as external constraints. The authors in [15] provided a different solution to ensure semantic compatibility of web services. However, the same arguments as for [14] regarding the absence of a high-level component model hold true.

With regard to the application architecture aware adaptation, Rainbow [16][17] is one of the most dominant and well-known frameworks. Rainbow uses invariants for the specification of constraints in its architecture description language. For each invariant a method for the adaptation of the system can be specified. The method is then executed whenever the invariant is violated. This approach however requires the knowledge of all component types at design-time, which is opposing our goal of an open system. Additionally, the developer has to implement the adaptation steps individually for every invariant. This imperative method for adaptation requires the component requiring the adaptation to have a view of the complete system and additionally introduces a big overhead at design-time as well as at run-time.

R-OSGi [18] takes advantage of the features developed for centralized module management in the OSGi platform, like, e.g., dynamic module loading and –unloading. It introduces a way to transparently use remote OSGi modules in an application while still preserving good performance. Issues like network disruptions or unresponsive components are mapped to events of unloaded modules and thus can be handled gracefully – a strength compared to many other platforms. However R-OSGi does not provide means to specify application architecture specific requirements. As long as modules are compatible with each other they will be linked. The module developer has to ensure the application architecture at the implementation level. Opposed to that, our approach proposes a high level description of application architectures through application templates that can be specified even after the required components have been developed.

III. The Core of the Existing DAiSI Platform

This section will introduce the foundations of the DAiSI platform, consisting of a dynamic adaptive component model, a domain architecture model and a decentralized configuration service.

As already briefly mentioned DAiSI components interact with each other through services. Each DAiSI component consists out of a set of component configurations. Each component configuration defines a set of required services and a set of provided services. Figure 1 shows a sketch of a DAiSI component with some explanatory comments for an athlete in the biathlon sports domain.

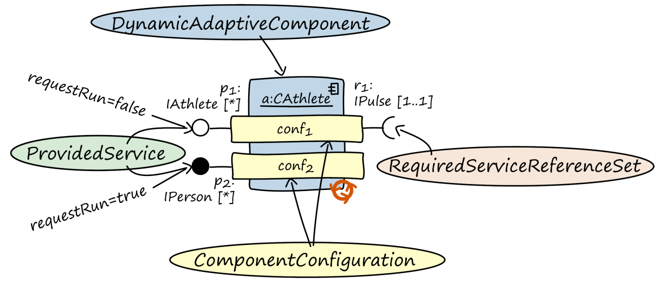


Figure 1. Notation for DAiSI components and corresponding concepts.

A component is depicted as a blue rectangle. Component configurations are bars that extend over the borders of the component and are depicted in yellow here. Associated to the component configurations are the provided and required services. The notation is similar to the UML lollipop notation [19] with full circles resembling provided, and semi circles representing required services. A filled circle indicates that the service is meant to be executed and thereby provided within the system, even if no other service requires its use.

Figure 1 shows the *CAthlete* component, consisting of two component configurations: $conf_1$ and $conf_2$. The first component configuration requires exactly one service variable r_1 of the *IPulse* interface. The second component configuration does not require any services to be able to provide its service p_2 of *IPerson*. The service can be used by any number of service users (the cardinality is specified as *). The other component configuration ($conf_1$) could provide the service p_1 of the type *IAthlete*, which could again be used by any number of users. The small orange circle with the three arrows in the lower right corner indicates that this component is self-organizing, i.e., it does not require a centralized configuration service to resolve its requirements and change its execution state.

Figure 2 shows the DAiSI component model as an UML class diagram [19]. The *DynamicAdaptiveComponent* class represents the component itself, represented as the light blue box in the notation example. It has three types of associations to the *ComponentConfiguration* class, namely *current*, *activatable*, and *contains*. The *contains* association resembles the non-empty set of all component configurations. It is ordered by quality from best to worst, with the best component configuration being the most desirable. Quality refers to the count of provided services, as well as the quality they are provided in. A subset of the contained are the *activatable* component configurations. These have their required services resolved and could be activated. An *active* component configuration produces its provided services. The *active* component configuration is represented by the *current* association in the component model, with the cardinality allowing one or zero current component configurations to be executed for a component.

The required services (represented by a semi-circle in the component notation in Figure 1) are represented by the *RequiredServiceReferenceSet* class. Every component configuration can declare any number of required services. The *resolved* association represents those that are resolved. Provided services (noted as full circles on the left hand side in Figure 1) are represented by the *ProvidedService* class. The flag

 ISBN: 978-1-61208-421-3

requestRun, represented by the full circle being filled with black in the component notation, indicates that the service should be activated, even if no other service requires its use. This is typically the case for services that provide graphical user interfaces or some functionality directly to the end user.

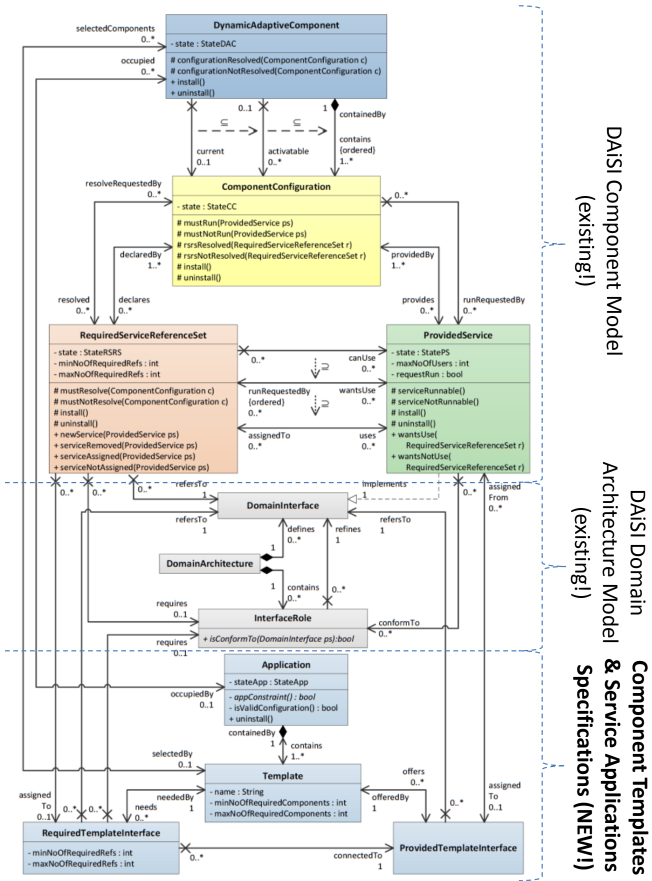


Figure 2. DAiSI component model, DAiSI domain architecture model, and the new additional template & application specifications.

The provided and required services, more precisely their respective classes in the component model, are associated with each other through three associations. The first association *canUse* represents the compatibility between two services. If a provided service can be bound to the service requirement of another class, these two are associated through a *canUse* association. A subset of the *canUse* association is *wantsUse*. At run-time it resembles a kind of reservation of a particular provided service by a required service reference set. After the connection is established and the provided service satisfies the requirement, they are part of the *uses* association, which represents the actual connections. All classes covered to this point implement a state machine to maintain the state of the DAiSI component. If you want to know more about the state machines and the configuration mechanism, please refer to [20].

To this point we have covered the basic building blocks of the DAiSI component model. Another already established part of DAiSI is the domain architecture model. The relation to the actual developed application becomes apparent if you consider the *DomainArchitecture* class. It defines any number of *DomainInterfaces*. These are the interfaces that define the provided and required services. Thus, every *ProvidedService* class implements a domain interface, while each required service reference set refers to exactly one.

There are numerous examples in which the role of a specific domain service has to be considered in order to establish the desired system configuration. For that reason the class *InterfaceRole* enables the specification of additional criteria for the conformance of provided and required services. An interface role references exactly one domain interface and may define additional requirements regarding that domain interface. A provided service only fulfills an interface role if it implements the domain interface and as well complies with the conditions defined in the interface role. Consequently a required service reference set not only requires compatibility of the domain interface, but also of the interface role to be able to use a provided service. For more information about the DAiSI domain architecture model and interface roles consider [21].

Beside the DAiSI component model and the DAiSI domain architecture model a decentralized dynamic configuration mechanism was also already established in the DAiSI platform. The set of services that implement the domain interface referred by the *RequiredServiceReferenceSet* is represented by *canUse*, as stated before. Note, this only guarantees a syntactically correct binding. Interface roles in addition provide a compatibility check with respect to a given common domain architecture. In [22][23] we have shown how this approach can be extended to guarantee behavior correct binding during run time, even in case of changes to the local and global state.

The *wantsUse* set holds references to those services for which a usage request has been placed by calling *wantsUse*. And the *uses* set contains references to those services, which are currently in use by the component or by *RequiredServiceReferenceSet*. Each time a new service becomes available in the system, the new service is added to all *canUse* sets, if the corresponding *RequiredServiceReferenceSet* refers to the same *DomainInterface* as the *ProvidedServices*. If there is a request for dependency resolution, usage requests are placed at the services in *canUse* by calling *wantsUse* and those service references are copied to the *wantsUse* set.

The management of these three associations—*canUse*, *wantsUse* and *uses*—between *RequiredServiceReferenceSets* and *ProvidedServices* is handled by DAiSI's decentralized dynamic configuration mechanism. This configuration mechanism relies on the state machines presented in more detail in [20] and sketched in the following paragraphs.

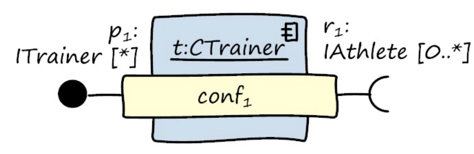


Figure 3. CTrainer component.

Assume a given component as shown in Figure 3. The component *t* of type *CTrainer* has one single configuration. It provides a service of type *ITrainer* to the environment, which can be used by an arbitrary number of other components. The component requires zero to any number of references to services of type *IAthlete*.

The boolean flag *requestRun* is true for the service provided. Hence, DAiSI has to run the component and provide the service within the dynamic adaptive system to other components and to users. As the component requires zero references to services of type *IAthlete*, DAiSI can run the component directly and thereby provides the component service to other components and users as shown in the sequence diagram in Figure 4.

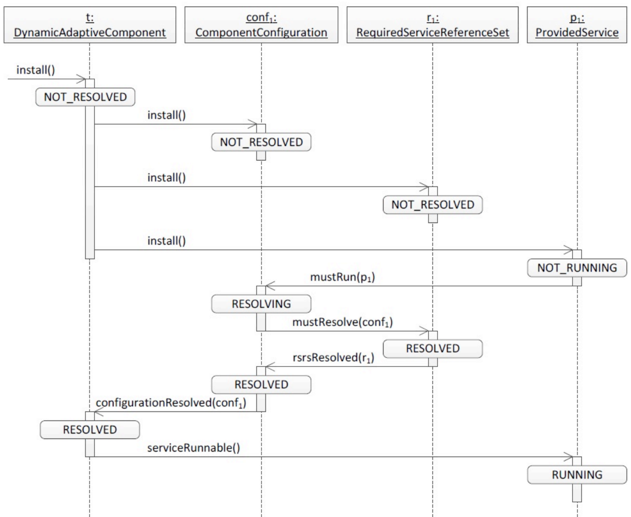


Figure 4. Sequence diagram showing the triggers and states of a stand alone DAiSI component.

Now assume two components: The *CAthlete* component, shown on the left hand side of Figure 5, requires zero or one reference to a service of type *IPulse*. The second component, *CPulse*, shown on the right hand side of Figure 5, provides a service of type *IPulse*. Note, this service can only be exclusively used by a single component. Figure 6 shows the states and triggers of the involved state machines in a sequence diagram for this example.

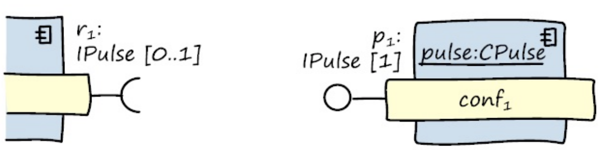


Figure 5. CAthlete and CPulse components.

Once the *CPulse* component is installed, DAiSI integrates the new service in the *canUse* relationship of the *RequiredServiceReferenceSet* r_1 of the component *CAthlete*. Then DAiSI informs the *CAthlete* component that a new service that can be used is available. DAiSI indicates that *CAthlete* wants to use this new service by adding this service in the set of services that *CAthlete* wants to use (set *wantsUse*).

Once the service runs, it is assigned to the *CAthlete* component, which can use the service from now on (added to the set uses of *CAthlete*).

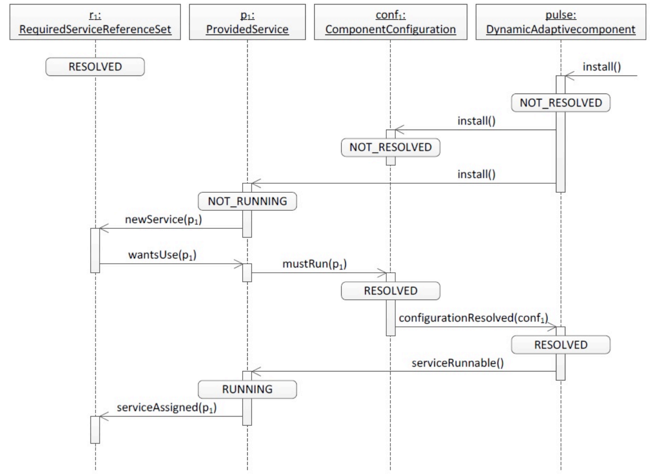


Figure 6. Inter-component configuration mechanism.

A detailed example of the presented configuration algorithms is presented in [20].

IV. Introduction of the Running Example and The Need to Control System Configurations

For this example, we assume that a self-organizing system is to be developed, which supports the training of biathletes, such as briefly described in the previous sections. In this particular case, the system is to provide the services described below.

First, a trainer is to be presented with an overview of his athletes' performance data, where data from at least one athlete should be displayed. For this purpose, it is assumed that the component presented in Figure 7 is available.

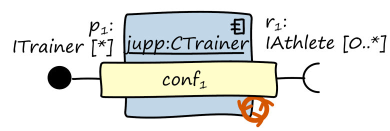


Figure 7. The trainer component available in the system.

The required functionality is provided by the service p_1, which implements the interface *ITrainer*. The service defines a dependency with services that implement the interface *IAthlete*. However, the service can also be run when an athlete system is not available in the system. The implementation of the trainer component would have to be adapted in order to meet the requirement that the trainer service can only be run when it has access to at least one athlete service. Moreover, the attribute *minNoOfRequiredRefs* of r_1 from Figure 7 would have to be set to 1. However, a component code cannot always be modified in this way. In addition, adapting it manually for the specific application purpose contradicts the original purpose of a component. The solution presented in the remainder of this section allows the application-specific specification of the minimum and maximum number of required references for *RequiredServiceReferenceSets* without having to adapt the component source code.

The individual athletes' performance data within the application are provided via the interface *IAthlete*. For the example, it is assumed that the component presented in Figure 8 is available.

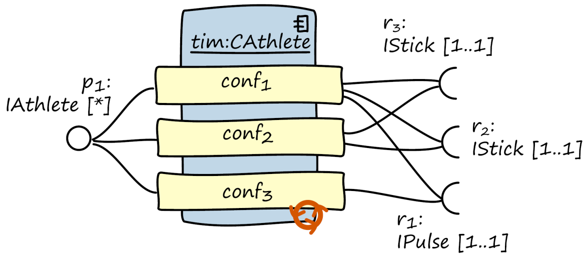


Figure 8. The CAthlete component.

The component defines three *ComponentConfigurations* with $conf_1$ specified as best configuration and $conf_3$ as worst. The $conf_3$ configuration can be activated if r_1 can be connected to a service that implements the interface *IPulse*. The $conf_2$ configuration can be activated, if r_2 and r_3 are each connected with a ski pole. The $conf_1$ configuration is activated if the dependencies of all three *RequiredServiceReferenceSets* can be resolved. In all three configurations, the component provides a service that implements the domain interface *IAthlete*. It defines a method *getPulse():int* to query the current pulse and also a method *getSkiingTechnique():String*, which returns the currently used skiing technique (double poling/diagonal technique). If the $conf_3$ configuration is active, the call *getSkiingTechnique* returns the value *null*. If, in contrast, the $conf_2$ configuration is active, the call *getPulse* returns the value -1.

For the example application, it is now assumed that the skiing technique is to be analyzed in particular, i.e., only the $conf_2$ *ComponentConfiguration* of the athlete component tim from Figure 8 is relevant. Even if one pulse service and two ski pole services are available, the $conf_1$ configuration should not be activated even though it is the best component according to the component specification. The framework presented so far, and described in Section III, does not provide the potential to influence the *ComponentConfiguration* of a component from an application-specific point. In this context, it is only possible to implement the component specifically to the application. In this section, expansions of the existing framework are described, which enable such an application-specific influence on the activation of component configurations.

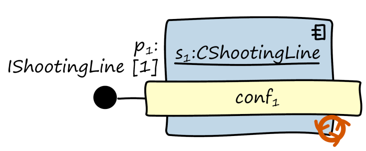


Figure 9. The CShootingLine component.

It should also be possible for the example system described here to allow shooting training. In this case, one shooting lane should be available for each athlete. In the system, each shooting lane should be represented by a service, each implementing the domain interface *IShootingLine*. One example of such a component is presented in Figure 9.

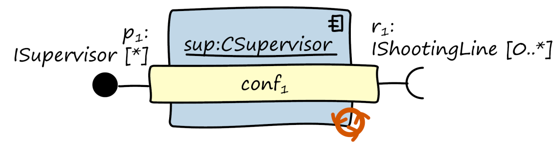


Figure 10. The structure of the component CSupervisor.

In this case, the service p_1 of the component also starts when there is no user in form of another component, as the flag *requestRun* is set (indicated by the shaded circle). However, for this example, the system should only allow shooting if a shooting supervisor is present. This is represented in the system by a service that implements the domain interface *ISupervisor*. The component presented in Figure 10 provides such a service.

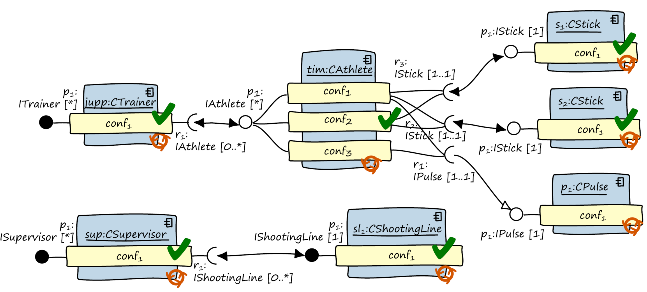


Figure 11. A system configuration that meets the requirements.

At this point, the most complex requirement placed on the system has an influence. The system must guarantee that exactly one shooting lane component is available for each athlete connected to the trainer component. This means that the number of those services used by the shooting supervisor component must be in agreement with the athlete components, which the trainer component accesses.

One system configuration that meets all criteria described above is presented in Figure 11. Here, a trainer component is connected with an athlete component, which in turn is connected to a left and a right ski pole. In addition, the application consists of a shooting supervisor component, which in turn is connected to a shooting lane component.

In the current DAiSI, such system configuration requirements cannot be specified and therefore cannot be guaranteed. Moreover, further requirements would be relevant for this application, such as: if a new athlete component is added to the system in the configuration described above, it should only be integrated into the application when a shooting lane component is available for this athlete. The application is also stopped, for example, when the athlete component from Figure 11 is only connected with one ski pole component.

DAiSI as described in Section III (without the new part for the specification of the component templates and application specifications) is not able to implement requirements relating to the application as a whole. For example, the better $conf_1$ configuration of the component "tim" from Figure 11 would be activated, although this is explicitly considered undesirable by the application developer.

V. Application Specific System Configuration

This section provides a short overview of the solution approach for the specification of valid system configuration requirements. One application configuration consists of a number of components, as well as connections between these components. Therefore, the primary task of DAiSI is to select the components that can be considered for a configuration conforming to the application architecture out of the number of all available components. In addition, the components must be connected in such a way that all specified requirements are met.

The solution presented below enables requirements specification with which components are considered for use within the application. On the other hand, the manner in which these components are to be connected with each other can be defined. Based on such a specification and number of components available, the framework developed will later be able to create an application architecture-conform configuration. Furthermore, the framework reconfigures the application automatically, if the requirements are no longer met. The solution approach is continuously based on a system of self-organizing components. However, the configuration is accomplished with the assistance of a central but light configuration unit.

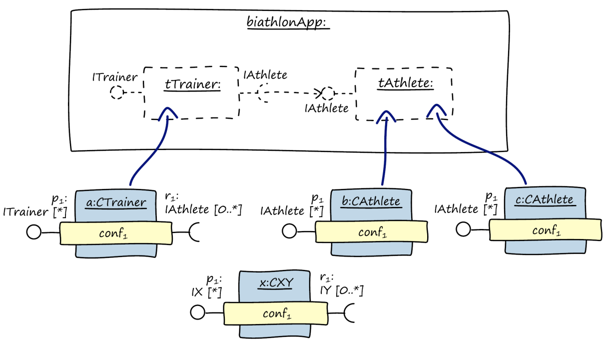


Figure 12. Suitable components for an application configuration.

The criteria for selection of suitable components for an application are defined with the assistance of so-called templates. An application specification consists of one or more of such templates. In this way, the biathlon application described above could, for instance, consist of a template for trainer components, and one for athlete components, one for shooting lane components, etc. For each of these templates, requirements can be stored that specify under which circumstances a component is compatible with a template. For example, constraints can be stored for an athlete template, which specifies that only such components that provide a service that implements the domain interface *IAthlete* are compatible. The framework ensures that for the runtime, only components matching the outline are allocated to the template. From then on, a template will be represented by a rectangle with dashed lines. Requirements related to required and provided component services are represented visually by circles and semi-circles with dashed lines (described in detail below). In Figure 12, two placeholders within an application template can be seen. One or two components can be allocated to the application, while one of the given components remains ignored, as it is not compatible.

The components selected must be connected with each other in the next step, in order to obtain an executable system. For this purpose, in addition to the templates, the links between templates are defined, and represented as dashed arrows (see Figure 12). They provide information on how the allocated components are to be connected with each other. In this way it is possible to define that each component allocated to the *tTrainer* template in Figure 12 must be connected with at least one component, which is allocated to the *tAthlete* template. Later during run time, the framework ensures that the requirements related to the links between the components are considered. Figure 13 shows one possible resulting system configuration.

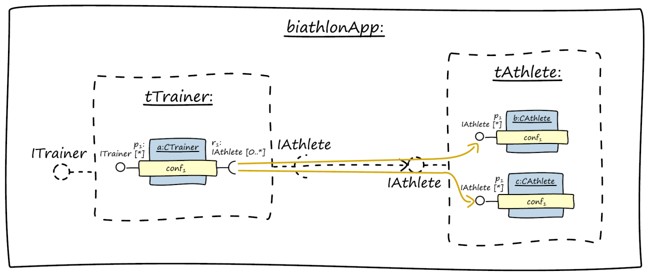


Figure 13. Generation of a valid configuration.

The following paragraphs present the requirements in detail, how they can be specified and how they are implemented in the framework.

VI. Specification of Application Requirements with Component Templates and Service Applications

The DAiSI platform is expanded to describe application-specific requirements for system configurations. These expansions represent the new parts of DAiSI in Figure 2, which are necessary to specify application-specific requirements for the system configurations, with the assistance of component templates and service application specifications.

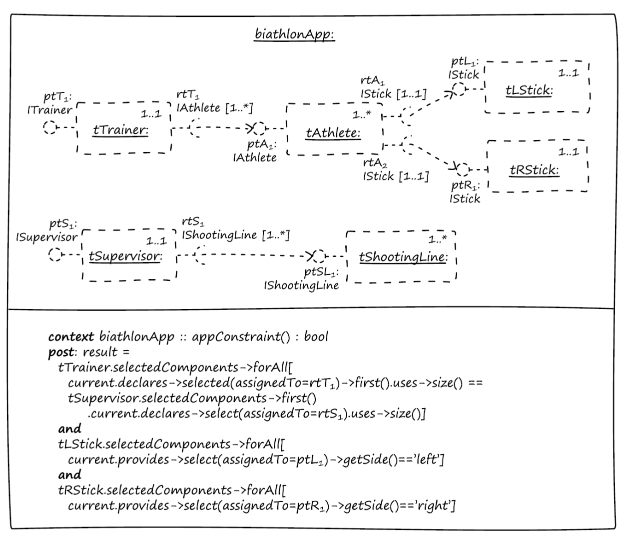


Figure 14. Graphical and textual application specification.

In order to be able to present application specifications in a concise manner, a uniform notation for the specification elements introduced above is defined for the remainder of this work. Figure 14 shows a possible application specification for a biathlon application.

An application is represented as rectangle with the name of the application noted at the top. Each Template is represented as a rectangle with dashed lines, which contains the name of the template. Within a template, the contents of the attributes *minNoOfRequiredComponents* and *maxNoOfRequiredComponents* are noted at the top right. A *ProvidedTemplateInterface* is represented as circle with dashed line, which is labeled with the name, as well as the referenced domain interface. *RequiredTemplateInterfaces* are represented correspondingly as semi-circles with dashed lines. They are also labeled with the referenced domain interface, the referenced interface role, if applicable, and with the name. Links between a *RequiredTemplateInterface* and a *ProvidedTemplateInterface* (*connectedTo*) are visualized with a dashed arrow. The predicate *appConstraint* specification is specified in a separate area under the templates.

VII. Requirement Conform Dynamic Adaptive Application Configuration

The aim of the framework is to create an application configuration, which meets all specified requirements. As soon as this is achieved, the applications' state machine transitions from NOT_RUNNING to RUNNING. In other words: if an application is in the state RUNNING, the application configuration created conforms to the application architecture.

This section describes how a valid application configuration can be generated automatically. The method suggested here follows a brute-force approach, which iteratively generates all possible configurations. It is sketeched in Figure 15 as pseudo code. While this is not optimal with regard to resources, it is sufficient to generate a valid system configuration. The focus of this paper is not the configuration algorithm, but the introduction of application templates.

```
1  boolean createValidConfiguration() {
2    while(possibleComponentAssignmentSets.hasNext()) {
3      possibleComponentAssignmentSets.next().realize();
4
5      while(possibleInterfaceAssignmentSets.hasNext()) {
6         possibleInterfaceAssignmentSet.next().realize();
7
8        while(possibleUsageSets.hasNext()) {
9          possibleUsageSets.next().realize();
10
11         if(isValidConfiguration()) {
12           return true;
13         }
14       }
15     }
16   }
17   return false;
18 }
```

Figure 15. createValidConfiguration() method, pseudo code listing.

Since a valid configuration, which meets the requirements can change at any time, in such a way that it no longer conforms to the application architecture, the application configuration is checked cyclical for conformance to the application architecture. Just as the configuration algorithm offers room for improvements with regard to performance, the same holds true for the cyclical application architecture conformance checks.

As soon as the configuration no longer meets the defined application architecture-specific requirements, and therefore the predicate *isValidConfiguration* is evaluated as *false*, the applications' state machine changes back to the state NOT_RUNNING.

The main task of the configuration process is to use a number of components to create an application configuration meeting all the requirements. For this purpose, the framework initially creates a configuration that meets all structural requirements. This configuration is executed and the services commence. It is in the next step a check is made to ensure that the service state requirements are met, since these requirements can only be confirmed when these services are running. If the requirements are not met, a new structurally compatible configuration must be created.

The algorithm is divided into two parts: one part creates an application configuration (lines 2-9 in Figure 15) and the other parts checks the configuration for conformity with the requirements lines (11-12 in Figure 15). Creating a configuration requires three steps. Firstly, selecting the components, then the *ProvidedService*- and *RequiredServiceReferenceSets* must each be allocated to a *ProvidedTemplateInterface* and *RequiredTemplateInterface*, respectively. Therefore, the two *assignedTo* quantities must be defined. Finally, the *uses* set must be determined for each *RquiredServiceReferenceSet*.

The initial situation of the configuration process is a set of available components. A selection must be made to obtain an application configuration. To accomplish this, assignment of the *selectedComponents* set is created for each template, with the component static properties already being considered. The application calculates the set of all possible assignment combinations and makes them available via an iterator (*possibleComponentAssignmentSets* from Figure 15), based on the components available and the application specification, the method *realize* implements the specific assignment.

For clarification, study an application specification with two templates as presented in Figure 16. It is also assumed that five components are available in the system.

In this example, the components *a* and *b* can be allocated to the *tTrainer* template and just one component must be allocated to the template in order to fulfil the application requirements. Both components provide a service that implements the *ITrainer* domain interface and define a *RequiredServiceReferenceSet* that references the *IAthlete* domain interface. Only component *d* can be allocated to the *tAthlete* template since this component is the only one that meets the template structural requirements. A total of two components are available for the *tLStick* and *tRStick* templates and exactly one component must be allocated to each of these two templates, in order to be able to meet the application requirements. This results in a number of possible allocations of components to templates. The configuration algorithm makes a selection, which is then implemented by the framework. In the next step, *ProvidedServices* is allocated to *ProvidedTemplateInterfaces* and *RequiredServiceReferenceSets* to *RequiredTemplateInterfaces*.

 ISBN: 978-1-61208-421-3

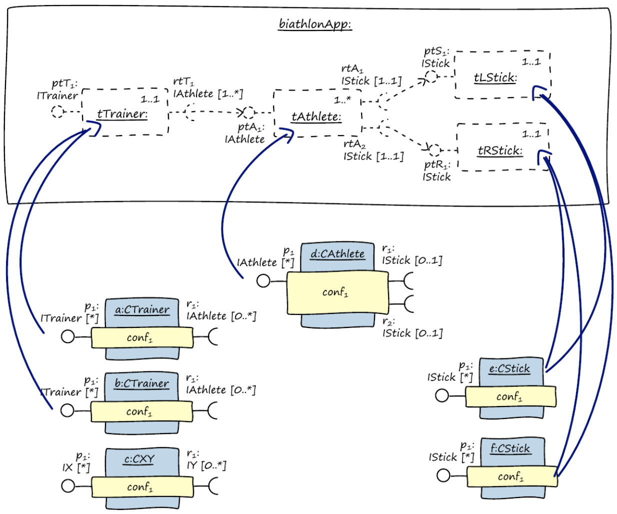


Figure 16. Allocation of components to templates.

ProvidedServices of a component could fit to several *ProvidedTemplateInterfaces*. Since *ProvidedServices* must be allocated to *ProvidedTemplateInterfaces* during run time, the framework must make a decision here. The same applies to *RequiredServiceReferenceSets* and *RequiredTemplate-Interfaces*. For example, the *RequiredTemplateInterfaces* of the *tAthlete* template in Figure 16, do not reference any interface roles but only the *IStick* domain interface as presented in Figure .

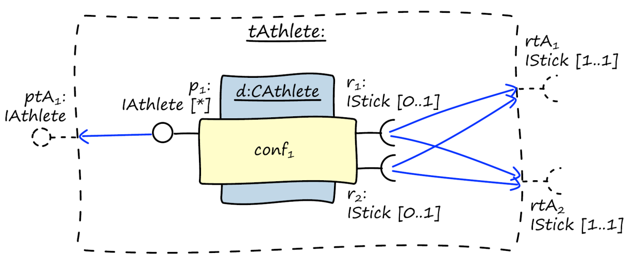


Figure 17. Allocation of component interfaces to template interfaces.

In this example, the *RequiredServiceReferenceSet* r_1 can be allocated to *RequiredTemplateInterface* rtA_1 as well as rtA_2. The same applies to RequiredServiceReferenceSet r_2.

Within the algorithm in Figure 15, all possible allocations, which result from the allocation of components to templates in the previous step are now iterated. In the component model, the allocation between *RequiredServiceReferenceSet* and *RequiredTemplateInterface*, and between *ProvidedService* and *ProvidedTemplateInterface* are represented by the *assignedTo* association. The possibilities are iterated with the *possibleInterfaceAssignmentSets* iterator (lines 5+6). The returned assignment is then implemented by calling *realize*. In the next step, the *uses set* is assigned to the *RequiredServiceReferenceSets* of the components, which were allocated previously to the *selectedComponents* quantity. The last step for the generation of the configuration algorithm consists of creating the *use* relations between the components.

The goal is assignment of the *uses* set for each *RequiredServiceReferenceSet* of all components included in the application, so that the requirements of the application specification can be met.

It often happens that there are several possibilities for the assignment of this set. The following situation is considered for illustration purposes (see Figure 15). In this case, the use of the provided service for both athlete components is considered for the *RequiredServiceReferenceSet* r_1. In this case, the empty quantity would not be an invalid assignment since the value 1 is specified for the attribute *minNoOfRequiredRefs* of the component. In the algorithm in Figure 15, these possible assignments are iterated with *possibleUsageSets* iterator, in order to create valid application configurations.

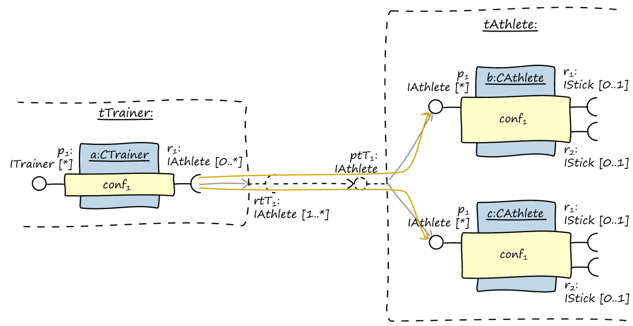


Figure 15. Example for possible assignments of the quantity *uses*.

After making a component selection, subsequently allocating the services and then assigning the *uses* set of all *RequiredServiceReferenceSets*, a running configuration is created automatically. For this purpose, the self-configuring components are informed at each stage if they are part of the application, to which template they should allocate themselves, to which template interfaces their services and *RequiredServicesReferenceSets* should be allocated and with which service they should connect. The individual iterators of the algorithms are realized for individual components.

After creating a configuration with the procedure described above, the remaining applications of the application specification can now also be checked for conformity. The predicate *isValidConfiguration* must now be evaluated. Only if this predicate is evaluated to *true*, the application changes its state to RUNNING. Otherwise, a new configuration must be created. The algorithm presented here is only a sketch of the procedure for creating a configuration, which conforms to the defined application architecture-specific requirements. Other algorithms are possible and can be found in [24].

VIII. Conclusion and Future Work

In this paper we introduced a major extension of our dynamic adaptive system infrastructure called DAiSI. DAiSI enables applications to adapt themselves automatically during run time. It is able to integrate new components during and handle the loss of components by reconfiguration. In the former version, DAiSI tried to find a configuration, which is optimal for each individual component. As this may lead to applications where each component is running in its optimal configuration, but where the application as a whole does not meet the requirements, we presented an extension of DAiSI, which enables the specification of application-specific requirements on the one hand, and its automatic realization during run time on the other.

Our concept introduced so called templates, which define fitting-criteria for component instances. Furthermore, the concept enables users to specify requirements regarding connections between components. Our infrastructure is able to interpret this specification, and realize a suitable application configuration based on available components in the system. One of the major characteristics of our approach is, that during design-time no knowledge about existing components and their instances is required. The match of components to templates is performed automatically during runtime based on provided/required interfaces, interface roles and predicates.

In the future, we will further extend our concept and our implementation by providing more specification capabilities regarding component selection and component interconnection. There is for example a possibility missing to specify an order on available components to enforce the use of, e.g., the best three components. Furthermore, there still exists potential for improvements of our prototypical implementation.

However, the extension presented in this paper provides a sustainable concept towards the realization of decentralized, dynamic adaptive systems, while satisfying application-specific requirements, which has been implemented as a proof of concept.

REFERENCES

[1] L. Northrop, et al., "Ultra-large-scale systems—the software challenge of the future," Software Engineering Institute, Carnegie Mellon, Tech. Rep., June 2006.

[2] D. Leffingwell and D. Widrig, "Managing Software Requirements: A Unified Approach," Addison-Wesley Professional, 2003.

[3] J. Kramer and J. Magee, "Self-managed systems: an architectural challenge," in FOSE '07 Future of Software Engineering, IEEE Computer Society, Washington DC, USA, 2007, pp. 259–268.

[4] B. H. C. Cheng, R. Lemos, H. Giese, P. Inverardi, and J. Magee, "Software engineering for self-adaptive systems: A second research roadmap," in Software Engineering for Self-Adaptive Systems II, Springer, Heidelberg, 2013, pp. 1–26.

[5] J. Kramer and J. Magee, "A rigorous architectural approach to adaptive software engineering," Journal of Computer Science and Technology, vol. 24, no. 2, 2009, pp. 183–188.

[6] D. Niebuhr, H. Klus, M. Anastasopoulos, J. Koch, O. Weiß, and A. Rausch, "DAiSI—dynamic adaptive system infrastructure," Technical Report Fraunhofer IESE, 2007.

[7] C. Szyperski, "Component Software," Addison Wesley Publishing Company, 2002.

[8] M. P. Papazoglou, "Service-oriented computing: concepts, characteristics and directions," in Proceedings of the 4th International Conference on Web Information Systems Engineering (WISE 2003). 10-12 December, Rome, Italy: IEEE Computer Society Press, 2003, pp. 3–12.

[9] J. Magee, J. Kramer, and M. Sloman, "Constructing distributed systems in conic," in IEEE Transactions on Software Engineering vol. 15, no. 6, 1989, pp. 663–675.

[10] J. Kramer, "Configuration programming: a framework for the development of distributable systems," in Proceedings of IEEE International Conference on Computer Systems and Software Engineering (COMPEURO 90), Tel-Aviv, Israel, 1990, pp. 374–384.

[11] J. Kramer, J. Magee, M. Sloman, and N. Dulay, "Configuring objectbased distributed programs in rex," Software Engineering Journal, vol. 7, no. 2, 1992, pp. 139–149.

[12] R. R. Aschoff and A. Zisman, "Proactive adaptation of service composition," in: H. A. Müller, L. Baresi (Eds.): Proceedings of the 7th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS'12): Zürich, Switzerland, June 4-5, 2012. Los Alamitos, California: IEEE Computer Society Press, 2012, pp. 1–10.

[13] A. Rasche and A. Polze, "Configuration and dynamic reconfiguration of component-based applications with microsoft .NET," in Proceedings of the 6th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC 2003). 14-16 May 2003, Hakodate, Hokkaido, Japan: IEEE Computer Society Press, 2003. ISBN 0-7695-1928-8, pp. 164–171.

[14] T. Kawamura, J.-A. De Blasio, T. Hasegawa, M. Paolucci, and K. Sycara, "Public deployment of semantic service matchmaker with uddi business registry," in The Semantic Web ISWC 2004, ser. Lecture Notes in Computer Science, S. McIlraith, D. Plexousakis, and F. van Harmelen, Eds. Springer Berlin Heidelberg, 2004, vol. 3298, pp. 752–766.

[15] T. Haselwanter, P. Kotinurmi, M. Moran, T. Vitvar, and M. Zaremba, "Wsmx: A semantic service oriented middleware for b2b integration," in International Conference on Service-Oriented Computing. Springer, 2006, pp. 4–7.

[16] D. Garlan, S. Cheng, A. Huang, B. Schmerl, and P. Steenkiste, "Rainbow: Architecture-based self-adaptation with reusable infrastructure," in Computer 37, 2004, No. 10, pp. 46–54.

[17] S. Cheng, "Rainbow: Cost-effective software architecture-based self- adaptation," Pittsburgh, Carnegie Mellon University, School of Computer Science, Dissertation, 2008.

[18] J. Rellermeyer, G. Alonso, and T. Roscoe, "R-Osgi: distributed applications through software modularization," in Proceedings of the ACM/IFIP/USENIX 2007 Conference on Middleware (Middleware '07), Newport Beach, California, 2007, pp. 1–20.

[19] OMG, OMG Unified Modeling Language (OMG UML), Superstructure, Version 2.4.1, Object Management Group Std., Rev. 2.4.1, August 2011. [Online, retrieved: 06, 2015], available from: http://www.omg.org/spec/UML/2.4.1

[20] H. Klus, A. Rausch, and D. Herrling, "DAiSI-dynamic adaptive system infrastructure: component model and decentralized configuration mechanism," in International Journal On Advances in Intelligent Systems, vol. 7, no. 3 & 4, 2014, pp. 595–608.

[21] H. Klus, A. Rausch, and D. Herrling, "Interface roles for dynamic adaptive systems," in Proceedings of ADAPTIVE 2015, The Seventh International Conference on Adaptive and Self-Adaptive Systems and Applications, 2015, pp. 80–84.

[22] D. Niebuhr and A. Rausch, "Guaranteeing correctness of component bindings in dynamic adaptive systems based on run-time testing," in Proceedings of the 4th Workshop on Services Integration in Pervasive Environments (SIPE 09) at the International Conference on Pervasive Services 2009 (ICSP 2009), 2009, pp. 7–12.

[23] D. Niebuhr, "Dependable dynamic adaptive systems: approach, model, and infrastructure," Clausthal-Zellerfeld, Technische Universität Clausthal, Department of Informatics, Dissertation, 2010.

[24] H. Klus, "Anwendungsarchitektur-konforme konfiguration selbstorganisierender softwaresysteme," translates to: Application architecture conform configuration of self-organizing softwaresystems, Clausthal-Zellerfeld, Technische Universität Clausthal, Department of Informatics, Dissertation, 2013.

Touch User Interface Evaluation for E-book Systems:

A Usability Experiment on Zinio and MagV

Yang-Cheng Lin
Department of Arts and Design
National Dong Hwa University
Hualien, Taiwan
e-mail: lyc0914@mail.ndhu.edu.tw

Chung-Hsing Yeh
Faculty of Information Technology
Monash University
Clayton, Victoria, Australia
e-mail: ChungHsing.Yeh@monash.edu

Abstract**—In this paper, we conduct a usability experiment to evaluate two popular e-book systems (Zinio and MagV) using a touchscreen device (iPad). In the experiment, we ask 24 experienced Internet users to perform 10 tasks conducted by five interface design experts. We then collect the quantitative and qualitative data, including the operation time, error frequency, and subjective satisfaction, to explore the usability problems of the touch interface design for e-book readers. The experimental results show that the interface design of Zinio is more user-friendly and efficient as perceived by the participants. In addition, although the touch interface is effective for the participants, the operational performance could be compromised by the poor interface design of the e-book system. The research outcomes provide useful references for interface designers in designing user-friendly interfaces of e-books.**

Keywords-touch interface; e-book; usability; iPad.

I. INTRODUCTION

As both 'ambient intelligence' and 'ubiquitous computing' grow and mature, mobile devices or handy devices have become some of the most desired and popular commercial products, including smart phones, e-book readers, ultrabooks (laptops), and tablets [1]. According to various marketing reports [2][3], the number of mobile device users is with 10%-30% annual growth. This is a fantastic outcome for manufacturers and product designers. Additionally, a trend has been noted that the development of e-books is gradually changing the reading habits of mobile users and the manner in which they retrieve information [4]. For example, in the United States, around one in four of all book-buyers purchase at least one e-book each month. Moreover, 31% of new books purchased are e-books, and 15% of the dollars spent on these books are for e-books [5]. However, can e-books meet users' demands for reading? Is reading an e-book the same as reading a physical book [6]? Can the readers adapt themselves to the differences in the touch interface of e-books?

To address these issues, we conduct a usability experiment [7] on e-books in this study. For our experiment, we choose the iPad with a touchscreen as the input device. In order to collect additional data for further analysis, we choose two popular e-book systems (apps): Zinio and MagV, due to their popularity and cross-platform attributes [1], which will be described in the subsequent sections.

In subsequent sections, we first present the usability experiment for touch interfaces, including the 10 experimental tasks. We then collect the quantitative and qualitative data based on the usability experiment, including the operation time, error frequency, and subjective satisfaction. Finally, the participants' recommendations and conclusions are given about the e-book systems.

II. USABILITY EXPERIMENT FOR TOUCH INTERFACES

The tablet is suitable and portable to read the text without zooming in and navigating around to have it perfectly positioned for reading [8]. In addition, the top tablet market share is iPad (26.9% in 2014 according to the International Data Corporation (IDC) report) [3]. As such, we choose the iPad as the experimental device, with two e-book systems. The Zinio system captures over 60% market share in digital magazine circulation, and it is a polished platform for digital magazine reading, which has apps for iOS, Android, Windows, and a desktop reader for Mac and PC. However, MagV is one of the biggest online Chinese bookstores in Asia, particularly in Taiwan, Hong Kong, and China. The Zinio and MagV e-book systems have different interface designs and different ways of navigating and reading e-books (as shown in Figure 1). Therefore, we can examine the usability of the interface design of these e-book systems and their different learning performances in reading e-books.

There are 29 participants, including five design experts and 24 experienced Internet users, involved in the experimental study. The 5 experts, with at least 10 years of interface design experience and the relevant analytical experiences of human-computer interface, form a focus group [7][9] to conduct 10 experimental tasks (as shown in Table I). The other 24 participants (with the average age of 22.8) are asked to test two e-book systems (Zinio and MagV) by using a touchscreen to perform 10 experimental tasks, whose results are used as a basis to conduct quantitative and qualitative analyses.

We record the entire experimental sessions on video while the participants are engaging in the experimental tasks. After the usability assessment, a semistructured questionnaire is used to collect information pertaining to the

 ISBN: 978-1-61208-421-3

'subjective satisfaction' of reading e-books. The procedure of this experiment involves the following steps:

(a) Zinio (b) MagV

Figure 1. Two e-book systems (apps): Zinio and MagV.

TABLE I. DESCRIPTION OF 10 EXPERIMENTAL TASKS

Task No.	Task Description	Corresponding Functions
T1	Please find the *AA* article on the Content Page of the *BB* book, and point out the exact position.	Scale Drag Page
T2	Please find the first page of the *AA* article, and point out the exact position.	Content Page Drag and Scroll Up/Down Page
T3	Please find the *CC* title in the *AA* article, and point out the exact position.	Drag and Scroll Up/Down Page
T4	Please read the *DD* point of suggestions aloud in the *CC* paragraph, and answer the *EE* question.	Scale Drag Page
T5	Please return to the initial model, and choose another *FF* book.	Content Page Choose Book
T6	Please browse the Content Page of the *FF* book, find the *GG* title, and point out the exact position.	Scale Drag Page
T7	Please find the first page of the *HH* article, which contains the *GG* title.	Content Page Drag and Scroll Up/Down Page
T8	Please answer the *II* question aloud in the *JJ* paragraph of the *HH* article.	Scale Drag Page Drag and Scroll Up/Down Page
T9	Please return to the *BB* book from the *FF* book, and find the first page of the *AA* article.	Content Page Choose Book Memory Retrieval
T10	Please find the *KK* paragraph in the *AA* article, and answer the *LL* question.	Scale Up/Down Page

* Each symbol *AA*, *BB*,…, *LL* represents a specific text individually.

Step 1: Conduct 10 experimental tasks by the five interface design experts.

Step 2: Perform the 10 experimental tasks by the 24 participants who test the two e-book systems by using the touchscreen. These participants are randomly assigned to test the Zinio or MagV e-book system. After the test, they take a short break (about 10 to 15 minutes), and then test the other system (MagV or Zinio), depending on which is taken first, until the experimental tasks are completed.

Step 3: Record the entire experimental sessions on video, including the operation time and errors occurred during the 10 experimental tasks.

Step 4: Distribute the semistructured questionnaire to collect quantitative (i.e., the subjective satisfaction of the participants) and qualitative data (i.e., the participants' suggestions regarding the e-book systems).

Step 5: Analyze the numerical data, including the operation time, error frequency, and subjective satisfaction.

III. ANALYSIS AND RESULT

In this section, we present the result of the quantitative and qualitative analysis, including the operation time, error frequency, and subjective satisfaction.

A. Operation Time

Table II shows the average operation time and error frequency of the 24 participants when they perform the 10 experimental tasks. In Table II, T1 toT10 in the first column represent Task 1 to Task 10, respectively. Table II shows that the operation time of the Zinio e-book system (229.02 seconds) is faster than the MagV system (261.21 seconds). The result suggests that the Zinio e-book system is easier to operate compared with MagV.

TABLE II. RESULT OF OPERATION TIME AND ERROR FREQUENCY

	Operation Time		Error Frequency	
	Zinio	MagV	Zinio	MagV
T1	12.42	24.65	0.17	0.33
T2	21.22	28.42	0.13	0.33
T3	16.22	5.88	0.04	0.00
T4	8.17	3.33	0.00	0.00
T5	22.64	22.78	0.33	0.25
T6	8.63	8.16	0.00	0.04
T7	43.98	104.94	0.79	0.92
T8	28.83	37.44	0.08	0.13
T9	37.70	19.63	0.29	0.13
T10	29.21	5.98	0.04	0.04
Total	**229.02**	**261.21**	**1.87**	**2.17**

For a further discussion, there is a significant difference when the participants perform T7 on the Zinio system (43.98 seconds) and the MagV system (104.94 seconds). The description of T7 is '*Please find the first page of the HH article, which contains the GG title*', and their corresponding operating functions are 'Content Page', 'Drag and Scroll', and/or 'Up/Down Page' (please refer to Table I). It indicates that when the participants perform T7 by using 'Content Page', 'Drag and Scroll', and/or 'Up/Down Page' functions, the operation performance on the Zinio system is significantly different as compared to the MagV system. In other words, the operation performance of an e-book system could depend largely on its user interface design.

B. *Error Frequency*

An error is defined as any action that does not reach the desired goal [7]. In this study, we count the number of such actions made by the participants while performing some specific tasks. The error frequency is the average of errors made by the participants in a specific task. From Table II, the error frequency of the Zinio e-book system (1.87 times) is less than the MagV system (2.17 times) for completing all the tasks. The result indicates that the Zinio e-book system results in fewer errors, thus having a faster operation time.

To further compare the two systems, we perform the correlation analysis to examine the relationship of operation time and error frequency. The result shows that there is a high positive correlation on the Zinio e-book system (Pearson's r=0.736, p=0.015), while the MagV e-book system has a very strong positive correlation (Pearson's r=0.951, p=0.001). The result reveals that the participants may make few errors while using a system, but the critical issue is how fast the users can recover from the errors [10].

C. *Subjective Satisfaction*

After the 24 participants complete the 10 experimental tasks, we distribute the semistructured questionnaire to collect quantitative and qualitative data. The participants are asked to assess two questions about the interface design (of the e-book system): 'S1: how do you feel regarding the visual and aesthetical style of the interface design?' and 'S2: how do you feel regarding the operation (operating efficiency) of the interface design?' A 5-point scale is adopted in the semistructured questionnaire, ranging from 1 (the lowest satisfaction) to 5 (the highest satisfaction). Figure 2 shows the result of the participants' subjective satisfaction (S1 and S2).

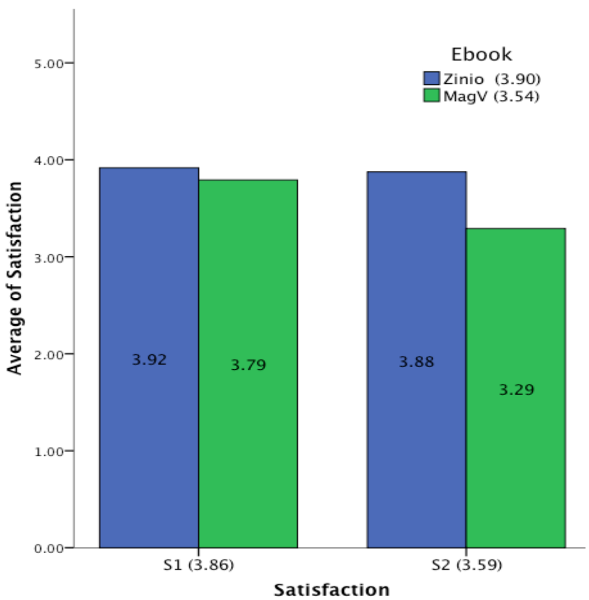


Figure 2. Result of the subjective satisfaction.

The result of S1 (i.e., aesthetics) shows that there is a slight difference between the Zinio system (3.92) and the MagV system (3.79), while S2 (i.e., operation) has a significant difference between them (the Zinio system being 3.88 and the MagV system being 3.29). We further perform the t-test to examine whether a statistical significant difference exists or not between these two e-book systems on S1 and S2, respectively. The result shows that the t value of S1 is 0.54 (p= 0.59> 0.05) and the t value of S2 is 2.30 (p= 0.03< 0.05). The result indicates that the Zinio interface design is more user-friendly and efficient (S2) as perceived by the participants, although the visual and aesthetical style (S1) of Zinio is identical to that of MagV. This is also reflected by the operation time and the error frequency.

Furthermore, the semistructured questionnaire is also used to collect the qualitative data. The participants are asked to write down their opinions and suggestions regarding the two different e-book systems, as given in Table III.

TABLE III. PARTICIPANTS' RECOMMENDATIONS

Zinio E-book System	**MagV E-book System**
• It is easy to read e-books with iPad, but lack of the feeling about real turning pages. • The page number should be consistent with the 'Scroll' page. • It is difficult to find the 'Content Page', and should be improved with hints/tips. • How to turn to the next page is inconsistent and confusing (e.g. some to the right-hand side, and others to the left). • The book should be marked if users have read it, and the label of books should be more visible and larger. • The 'Scroll' function is unclear, and easy to make an error to turn to the next page.	• The page number should be consistent with the 'Scroll' page. • Users cannot jump onto a specific page, but scroll page-by-page. • The duration of scroll page is slower than the reality. • For the convenience, books should be categorized into proper sections. • The whole operation is not very smooth, and not easy to find the specific content. • How to turn to the next page is inconsistent and confusing (e.g. some to the right-hand side, and others to the left). • Resolution of display is too low/fuzzy, particularly when the page is scaled up.

According to Table III, there are some common problems with respect to the two different e-book systems, such as 'the page number should be consistent with the 'Scroll' page'. In addition, although the iPad with the touchscreen is intuitive for the participants reading e-books, most of the participants feel confused and make errors on 'how to turn to the next page' (because some e-books are to the right-hand side, and others are to the left). For further analysis, most of the participants prefer combining two or three options (e.g., 'Up/Down Page' with 'Drag and Scroll') to read an e-book. In addition, the participants also like to navigate the article (i.e., page by page) in a similar manner to when reading a real book. This result could be provided as a guideline for the designers to design an effective user interface that can best meet user requirements and satisfaction.

IV. CONCLUSION

In this paper, we have presented a usability experimental study on two e-book systems (i.e., Zinio and MagV) to

Copyright (c) IARIA, 2015. ISBN: 978-1-61208-421-3

address whether the touchscreen device is user-friendly with high subjective satisfaction for e-book readers. The experimental study has been conducted to examine if the different e-book systems have significant difference in terms of the operation time, error frequency, and subjective satisfaction. The result has demonstrated that the Zinio e-book system is easier to operate with few errors compared with the MagV system. Moreover, even though the touchscreen device is intuitive for the participants reading e-books, the participants still feel confused and make errors due to the poor interface design. Although two e-book systems are chosen as an illustration, the user-centered approach presented in this study is applicable to other kinds of e-book systems (e.g., FlipViewer) for addressing the usability issue.

ACKNOWLEDGMENTS

This research was supported by the Ministry of Science and Technology, Taiwan under Grants NSC102-2410-H-259-069 and MOST103-2420-H-259-015.

REFERENCES

[1] Y.-C. Lin, "User-centred evaluation to interface design of e-books," Lecture Notes in Computer Science, vol. 8936, 2015, pp. 220-226.

[2] eMarketer, "Smartphone users worldwide will total 1.75 billion in 2014," January, 2014. http://www.emarketer.com/Article/Smartphone-Users-Worldwide-Will-Total-175-Billion-2014/1010536

[3] IDC report, "Worldwide tablet market grows 11% in second quarter on shipments from a wide range of vendors," Jul, 2014. http://www.idc.com/getdoc.jsp?containerId=prUS25008314

[4] V. Liesaputra and I. H. Witten, "Realistic electronic books," International Journal of Human-Computer Studies, vol. 70, 2012, pp. 588-610.

[5] BEA, "Measuring the e-book market," The 2013 Books and Consumers Annual Report, May, 2014. http://www.publishersweekly.com/pw/by-topic/industry-news/bea/article/62579-bea-2014-measuring-the-e-book-market.html

[6] W. D. Woody , D. B. Daniel, and C. A. Baker, "E-books or textbooks: Students prefer textbooks," Computers and Education, vol. 55, 2010, pp. 945-948.

[7] J. Nielsen, Usability Engineering. Academic Press, United Kingdom, 1993.

[8] CNET, "Tablets buying guide. CNET's Buying Guide," November, 2013 by Xiomara Blanco and Eric Franklin. http://www.cnet.com/topics/tablets/buying-guide/

[9] Y.-C. Lin, C.-H. Yeh, and C.-C. Wei, "How will the use of graphics affect visual aesthetics? A user-centered approach for web page design," International Journal of Human-Computer Studies, vol. 71, 2013, pp. 217-227.

[10] R. Budiu and J. Nielsen, Usability of iPad Apps and Websites. Nielsen Norman Group, 2011.

Improvement on K-means and TSP Based Mobility Protocol of Wireless Sensor Network

Afef Ghabri[1], Monia Bellalouna[2], Walid Khaznaji[3]
[1]Laboratory RIADI-GDL, [23]Laboratory CRISTAL POLE GRIFT
National School of Computer Sciences (ENSI)
Manouba, Tunisia
e-mail: [1]afef2108@hotmail.com, [2]monia.bellalouna@ensi.rnu.tn, [3]mwkhaznaji@yahoo.fr

***Abstract*—Wireless sensor networks are widely used in different environments in order to execute diverse tasks and applications. In this paper, we will consider problems related to the fault tolerance issue and we will present a new version of K-means And Traveling Salesman Problem based mobility protocol which aims to provide not only better energy efficiency within the wireless network, but also better reliability compared with the conventional method based on K-means clustering and the approximate solution for Traveling Salesman Problem by using the simple local search algorithm "2-Opt". This problem is Non-deterministic Polynomial-time hard (NP-hard), so we propose the new approach that navigates the mobile sink to go through the cluster centers according to the optimized route by implementing the local search method "Tabu". Simulation results have demonstrated that the solution given by the Tabu heuristic outperforms the original solution of K-means and Traveling Salesman Problem based mobility protocol in terms of quality. Our goal is to propose a much more realistic model that provides less execution time than the conventional strategy and an effective improvement once the problem is disturbed by the breakdowns of some nodes.**

Keywords-Wireless sensor networks; fault-tolerance; failures; Tabu; quality; realistic.

I. INTRODUCTION

The technological advances carried out in the past years have allowed the development of new varieties of sensors, which can be configured and used for wireless communication in order to form autonomous networks. Indeed, the sensors can send data to all the nodes which are connected to the network. They are enormously employed in various environments in a random way to perform different monitoring tasks such as the delivery, the searching, the help in case of disasters and the target tracking [1]. Generally, a sensor network is primarily composed of several collector nodes that are scattered throughout a sensing field, a data processing center and sinks [2]. Each node, randomly distributed, is responsible for collecting obtained data from its coverage area. The sinks are particular nodes that must always be active and whose role is to route recovered data coming from sensor nodes. Their number depends on the network load and size. These collection points represent the interface between the sensing field and the data processing center that facilitates data recovery and ensures their treatment in order to extract for the user useful information [3]. Figure 1 shows this principle.

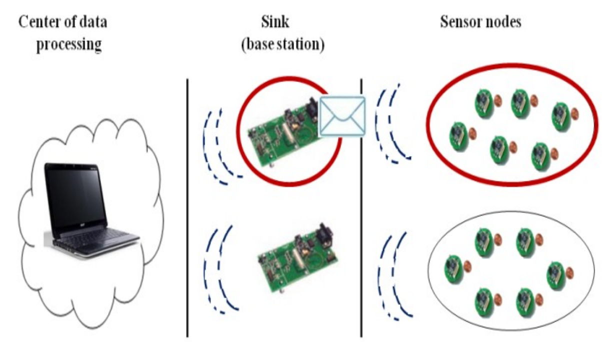


Figure 1. Wireless sensor network architecture

Generally, in the case of a large-scale network, data is periodically transmitted from source nodes to the sink up the tree- structure. Data routing is ensured by intermediate nodes that forward the messages until reaching the destination. Then, there exist different possible paths in order to derive information [4]. This process imposes energy conservation and network lifetime prolongation problems. It is thus necessary to find the shortest and optimal route in order to provide the data routing, within the energy constraints and the time of transferring packages [5]. In dealing with this problem, we noticed that it leads to a Combinatorial Optimization Problem (COP) [6]. In other words, several problems to be solved in the Wireless Sensor Network (WSN), such as the routing problem which returns to find a shorter path, or the minimization problem of energy, consumed by the sensors, taking into account several constraints as their displacement and their communications, can be reduced to optimization problems [7]. Unfortunately, the sensors are prone to many intrusions and failures, mainly due to physical destructions, depletion of batteries, communication link errors, malicious attacks and environmental interferences [8]. Thus, the sensor breakdown causes the loss of communication links that leads to a significant change in the overall network topology. This can affect the network connectivity and decrease its life [7]. Failures of sensor nodes should not affect the overall network performance. This is a problem of reliability or fault tolerance. Fault tolerance is the ability of a network to keep

 ISBN: 978-1-61208-421-3

its functionality without causing interruptions if a sensor node stops working. We can say that its objective is to avoid the total system flaw despite the existence of errors in a subset of its elementary components [9]. The studied problem is then reduced to a COP, where the number of data is a random variable instead of a deterministic value, and thereafter it was more interesting to consider it as being a Probabilistic Combinatorial Optimization Problem (PCOP) [10].

The purpose of this paper is to propose a new strategy that guarantees the obtaining of a solution in real time once the problem is disturbed by the failures of some nodes, by considering the fault tolerance as a PCOP through the K-means and Traveling Salesman Problem-based mobility (KAT-mobility) protocol [11]. The rest of the paper will be organized as follows: In Section II, we will present the PCOP and particularly the Probabilistic Traveling Salesman Problem (PTSP). Section 3 will describe its solving methods. A description of the KAT-mobility protocol principle will be shown in Section 4. In Section 5, we will propose our improved strategy. Section 6 presents the simulation results and analysis. Finally, we will conclude the paper in Section 7.

II. PROBABILISTIC COMBINATORIAL OPTIMIZATION PROBLEMS

In recent years, COPs are considered as a critical issue and a very interesting subject of research. The study of these problems became one of the most exciting and active areas in the field of discrete mathematics [12]. In fact, a COP is defined as follows: given n data, a set of configurations S and an objective function defined on S in $\Re$, the goal is to find a configuration s* that minimizes the function f.

$$\begin{aligned} f &: S \rightarrow \Re \\ s &\mapsto f(s) \end{aligned} \tag{1}$$

This deterministic model is inadequate with reality, where the number of data of the studied problem is frequently a random variable between 0 and n. Recently, a specific family of COPs, characterized by the fact that probabilistic elements are included explicitly in the problem definitions, has appeared and has been investigated [13]. For this reason, they were named PCOPs. There are many motivations for studying the effect of probabilistic elements inclusion in COPs: among them two are the most significant. The first one is the desire of formulating and analyzing models which are more convenient for real world problems, where randomness exists [12]. The second motivation is the possibility to analyze the robustness of optimal solutions for deterministic problems, considering the modification of the instances for which these problems are solved: disruption by the absence of certain data [14]. The PCOP is defined as follows: given a random number of data varying between 0 and n (0 when all the tasks are absent and n when all the tasks are present), a set of configurations S, an objective function defined on S in $\Re$ and a modification strategy which adapts the feasible solution through S to the new subset in order to obtain the realizable solution through the set of present tasks, the objective is to find a configuration s* which minimizes the functional of f according to this strategy.

$$E[f(s^*)] = \min(E[f(s)]) \tag{2}$$

The first studied problem in PCOPs, initially introduced by Jaillet [15], was the PTSP [16]. This approach has been then extended to other different problems and studies have continued in several domains, such as the Probabilistic Traveling Salesman Facility Location Problem (PTSFLP) [13], the probabilistic longest path problem [17] and the probabilistic minimum vertex covering problem [18]. The probabilistic approach was extended to combinatorial problems that are not defined on graphs, such as the Probabilistic Scheduling Problem (PSP) [19] and the Probabilistic Bin Packing Problem (PBPP) [20]. There are many interesting and important applications of PCOPs, particularly in the context of communication systems, strategic planning, job scheduling, etc. [21]. To deal with these problems, two strategies can be used: the re-optimization strategy [22] and the a priori strategy [13].

A. *Solving Strategies*

Frequently in applications, after having solved a particular instance of a given combinatorial optimization problem, we must solve repeatedly many copies of the same problem. These additional instances are generally simple variations of the original problem; however, they are sufficiently different to require an individual treatment [18]. The most natural approach used to address this kind of situation consists in solving in an optimal way the different potential copies. We call this strategy the "re-optimization strategy" [23]. However, it has many disadvantages and the most important one is the high cost. For example, if the considered COP is NP-hard [24], one could have to solve an exponential number of instances of a very hard problem. Moreover, in several applications it is required to find a solution to each new instance promptly, but one could not have the necessary computing or other resources to carry out such a task [13]. It is therefore necessary to adopt a different strategy. Rather than re-optimizing each successive exemplary, we can try to determine a priori solution of the initial problem that can be successively modified in a simple way to solve the following instances. We call this strategy, introduced by Jaillet [15][10] and which is less costly in terms of computations, the "a priori strategy" [6].

Different algorithms were implemented for PCOP resolution and have shown a satisfactory performance. Approximate methods have been used to solve large scale PCOPs. Among them, we mention the work of Bertsimas [25], in which he proposed and analyzed heuristics for PVRP. Many algorithms based on classic heuristics were also proposed for the PBPP, in the works of Bellalouna [6][22]. Metaheuristics were also employed to solve PCOPs, such as simulated annealing and Tabu search, implemented in the case of PTSP [22]. A Tabu Search was implemented by Gendreau et al. [26] for solving the vehicle routing problem with stochastic demands and customers [14].

 ISBN: 978-1-61208-421-3

Furthermore, exact methods have been used, such as branch & bound algorithm developed by Rosenow [27], to solve PCOPs and especially the probabilistic traveling salesman problem.

B. *Probabilistic Traveling Salesman Problem*

The PTSP is a variation of the standard Traveling Salesman Problem (TSP) [28]. It is essentially a TSP, in which the number of nodes that require being visited in each problem is a random variable [13]. In fact, for a given finite set of points, the TSP consists of finding a tour through the points of minimum total length [12], and in the probabilistic version of this problem, introduced for the first time by Jaillet [15]; only a subset of the nodes may be present in any given instance of it. The TSP can then be considered and treated as a special case of PTSP. Indeed, the principal difference between the two problems is that in TSP the probability of each visited node is 1, while in PTSP the probability of each visited node is between 0 and 1 [14].

In other words, consider a routing problem through a set of n known points. On any given instance of this problem, only a random subset S has to be visited. In many cases, we can not have resources to reoptimize the tour for every instance, and even if we had them, re-optimization may become too time consuming [13]. We wish to find a priori tour through all the points. On any given instance, the subset of present points will then be visited in the same order as they appear in the a priori tour, i.e., we simply skip the absent points in that problem instance [21]. The problem of finding such a priori tour of minimum expected length according to this skipping strategy is defined as a PTSP [12]. Figure 2 shows this principle.

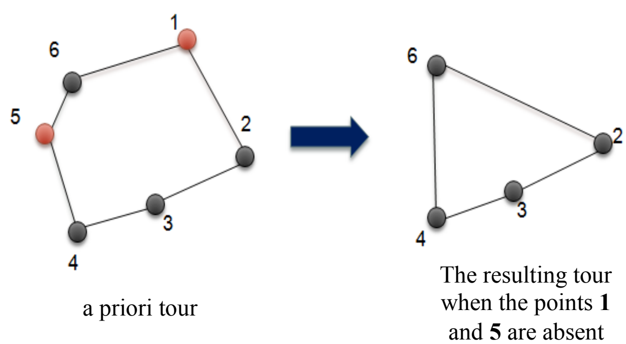


Figure 2. PTSP methodology

What distinguishes the PTSP from other problems is the probability distribution (law) that specifies the number and the identity of present points that require to be visited on any given problem instance. The PTSP can be formulated as follows: Consider a complete graph G = (V, E) on n nodes and a priori tour R through these points. If every possible subset of the node set V = {1, 2…, n} may or may not be present on any given instance of the optimization problem, for example, on any given day, the traveling salesman may have to visit only a subset S (S ⊆ V), then there are 2^n possible instances of the problem, i.e., 2^n is the number of all the possible subsets of V [29]. Let P(S) be the probability that instance S occurs. Given a method μ for updating a priori solution R to this optimization problem on G, μ will then produce for S, a feasible solution with the value (cost) $L_R(S)$. In fact, this solution is a tour through the subset S of nodes and $L_R(S)$ represents the length of that tour [15][29]. Then, given that we have already selected the updating strategy μ, the natural choice for the a priori solution is to select a tour through all potential nodes that minimizes the expected cost with the summation over all subsets of V [29].

$$\mathbb{E}\left[L_{(R,\mu)}\right] = \sum_{S \subset V} P(S) L_{(R,\mu)}(S) \quad (3)$$

Otherwise, each point i has a probability of presence p_i, independently of the others. We assume that d (i, j) is the distance between points i and j and we suppose that the a priori tour is R = (1, 2….., n-1, n, 1), then our problem is finding a tour through n nodes, which minimizes the expected length of a determined a priori PTSP tour R, denoted $\mathbb{E}\,[L_{(R,\mu)}]$ [14].

$$\mathbb{E}\left[L_{(R,\mu)}\right] = \sum_{i=1}^{n}\sum_{j=i+1}^{n} d_{ij} p_i p_j \prod_{k=i+1}^{j-1}(1-p_k) + \sum_{j=1}^{n}\sum_{i=1}^{j-1} d_{ji} p_i p_j \prod_{k=j+1}^{n}\prod_{k=1}^{i-1}(1-p_k) \quad (4)$$

If all the points have the same probability of presence ($p_i = p$, ∀ i), the expected length $\mathbb{E}\,[L_{(R,\mu)}]$ can be expressed using the following formula [6]:

$$\mathbb{E}\left[L_{(R,\mu)}\right] = p^2 \sum_{\tau}^{n-2} (1-p)^{\tau} L_{(R,\mu)^{\tau}} \quad (5)$$

Where

$$L_{(R,\mu)^{\tau}} = \sum_{i=1}^{n} d(i, R^{\tau}(i)) \quad (6)$$

R^{τ} comprises pgcd (n, τ+1) sub-tours. It consists in jumping τ points from the initial tour R. $R^{\tau}(i)$ represents the point after i along the permutation R^{τ}, so $L_{(R,\mu)}{}^{\tau}$ is the permutation length. We note that R^0 is the tour R and $L_{(R,\mu)}{}^{0}$ is the length of the tour R [14][15].

There are several methods for solving PTSPs. A wide variety of approximate and exact algorithms have been proposed for solving it. Exact methods can only solve relatively small problems. Concerning the approximate algorithms, a number of heuristics and metaheuristics have been effectively used in the case of large problems. Nowadays, approximate methods are considered as a very interesting subject for many researchers who still trying to find the best algorithm that gives a very good approximate solution in appropriate running time [14].

III. Solving Methods of PTSP

The PTSP was introduced by Jaillet [15], who studied some of its properties and proposed different heuristics and also an exact branch & bound algorithm to solve this problem [30]. It represents an optimization problem for which, there is no known algorithm that ensures the obtaining of an exact solution in polynomial time. The algorithms used for solving this problem can be classified into two categories: exact algorithms and heuristics. The exact algorithms permit finding an optimal solution, but they are characterized by a complexity that increases exponentially with the problem size (the number of cities). However, the approximation algorithms or heuristics permit the obtaining of good solutions, but can not guarantee their optimality. Thus, these methods approach the optimal solution in reasonable time. Among the most commonly used techniques for PTSP resolution, we can mention the local search algorithms 2-Opt and Tabu and the branch & bound technique [9].

A. 2-Opt Algorithm

The 2-Opt algorithm was proposed by Lin-Kernighan [31] in 1973, with a purpose of ensuring the obtained starting solution improvement [32]. Indeed, a possible transformation consists in removing 2 edges of the tour and recomposing another trajectory by reconnecting remaining segments in another way, through the use of new links and by reversing their course direction if possible (suppression of 2 edges then testing all the ways to reintegrate them). The 2-Opt algorithm provides an elementary transformation consisting in selecting two nonadjacent present edges in the Hamiltonian cycle and carrying out their exchange by two other edges in a manner to obtain a new cycle. One starts with a random tour and systematically tries to find a better tour by replacing two arcs of the tour by two other arcs [31]. Figure 3 illustrates this operation.

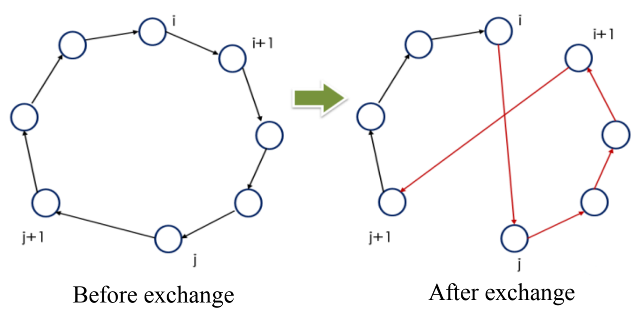


Figure 3. 2-Opt procedure

B. Tabu Algorithm

Tabu Search (TS) is a local search method combined with a set of techniques to avoid being trapped in a local minimum or cycle repetition. TS was mainly introduced by Glover [33] in 1986. This method has proved high efficiency for the resolution of difficult optimization problems. Indeed, from an initial solution in a set of local solutions, subsets of solutions that belong to this set neighborhood are generated. Through the evaluation function, we retain the solution which improves its value, selected among the set of neighboring solutions. The algorithm may accept sometimes solutions that do not always improve the current solution [26]. We implement a Tabu list containing the last visited solutions, which does not give the possibility for a solution, which is already found, to be accepted and stored in the Tabu list. Then, the choice of the next solution is carried out on a set of neighboring solutions outside of the elements of this list. When the list length is reached, each new selected solution replaces the oldest in the list. The list construction is based on the First In First Out (FIFO) principle. As a stopping criterion, one can for example fix a maximum number of iterations, or fix a limited time after which the search should stop.

TS is a local search method, and the structure of its basic algorithm is similar to that of simulated annealing, with the advantage of having a simplified parameter setting: the parameter setting consists initially in finding an indicative value of iterations during which the movements are prohibited. It will be necessary to choose a memorization strategy. However, the Tabu method requires a memory management, more and more intense, by putting complex memorizing strategies [33]. The effectiveness of the Tabu method provides its use in several classical COPs such as the TSP, the scheduling problem, the problem of vehicles, etc.

IV. KAT-MOBILITY PROTOCOL

Nakayama et al. [34] proposed the KAT-mobility protocol based on optimization methods of routing and aggregation [9]. It is a model of mobility of sinks that can effectively gather detected data employed in a WSN, even if some nodes are destroyed [35]. Indeed, the implementation of this mobility concept provides better energy management and increases the network lifetime. The system is composed of two modules: the clustering algorithm and the approximate solution for the TSP [34].

A. Clustering Algorithm

The KAT-mobility protocol is based on clustering and especially on the K-means approach [36] defined by McQueen [37] in 1967 and illustrated in Figure 4. The goal of this method is the division of the sensors set into k partitions, where each sensor will belong to the appropriate partition with the nearest average. The implementation of this algorithm is carried out by following these steps: the first step consists in choosing k that represents the number of groups to be created and then generating k initial centers randomly. The following step is to browse through all the nodes in order to assign them to the proper group whose center is the closest, on the basis of the computation of the distance between this center and each node. Then, we calculate the new centers associated with the new partition by seeking the average value of all the sensors in the respective group (centroid). Thereafter, we repeat the second step to perform the reassignment and the third step to update the average of each group. These last two steps will be then repeated until convergence to a stable partition: this convergence will be reached when there is no change. This algorithm convergence can be slow, so it is advised to add a stopping criterion, which is the case of KAT-mobility protocol. This process can be regarded as an optimization problem, as it aims to minimize the sum of the distances between group centers and the points located inside each group (the sum of the approximate errors). In fact, the group cost is estimated by the approximation error between the sensors and the center. This algorithm divides the set of nodes virtually into k clusters (C_1, C_2 … C_k) geographically close. We denote by n the nodes number in the network, usually n >> k. Let m_j (j = 1, 2 ... k) be a group center and x_i

(i = 1, 2 ... n) be a sensor, which is represented by a 2-dimensional vector (i.e., sensor position) [9]. d (x_i, m_j), which is indicated by the Euclidian distance between the group center and the sensor, represents the approximation error. The goal is to assign each sensor to a cluster C_j by reducing the total error of the clusters in order to reduce batteries consumption and communications energy [38]. The final objective is to assure the configuration of Cj such that the sum of approximate errors is minimized.

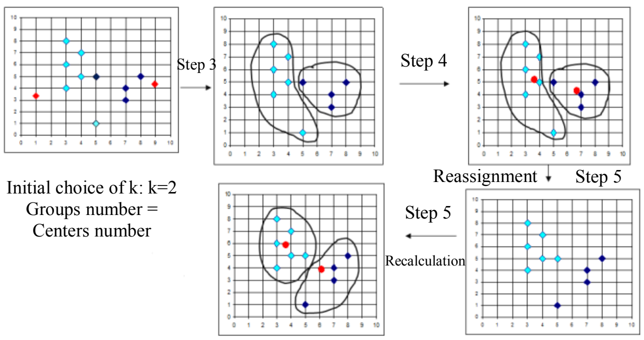


Figure 4. Operating principle of K-means algorithm

After the grouping of nodes, the K-means method reaches the mobile sink to make a course through the centers of groups, determined as anchor points, according to the trajectory of an optimized route [9]. This route is set as an approximate solution of the TSP.

B. Route Optimization

The second procedure of KAT-mobility protocol consists in optimizing the routing path of the collector (sink). Finding the best path for the mobile node is identical to the TSP. Then, a collector represents the traveling salesman and cluster centroids (centers) define cities. The route optimization of the mobile collector to visit once and only once every cluster centroid is equivalent to searching for the shortest trip of the traveling salesman in order to visit each city once [11]. Here, it is supposed that an administrator distributes nodes to supervise the targeted zone, and they are scattered at random locations and do not move afterwards. Indeed, this method reaches the mobile sink to make a course through the centers of groups according to the trajectory of an optimized path. The sink collects then the data coming from nodes on the level of the visited groups [9]. Figure 5 illustrates this principle [34].

The goal is to find the trajectory π which minimizes the tour lengths, where the initial position of the mobile sink is indicated by m_0. The quantity mentioned in (7) shows the tour length of a mobile collector that will be realized by visiting the centers in the specified order according to the permutation, while returning finally to the starting position.

$$\sum_{j=0}^{k-1} d\left(m_{\Pi(j)}, m_{\Pi(j+1)}\right) + d\left(m_{\Pi(k)}, m_{\Pi(0)}\right) \quad (7)$$

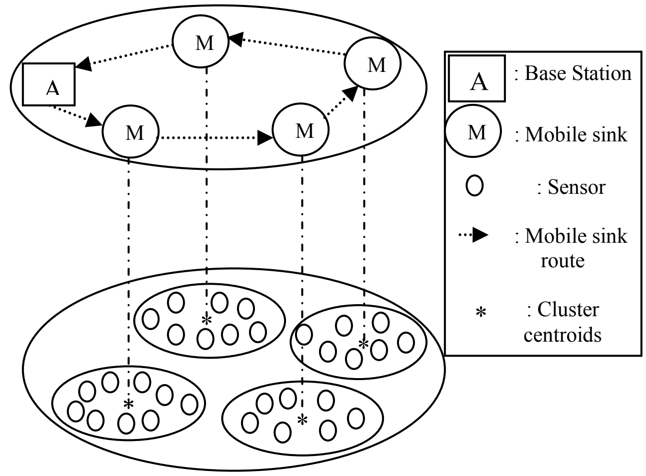


Figure 5. KAT-mobility protocol

The TSP belongs to the category of NP-hard problems and it represents an optimization problem, for which there does not exist a known algorithm that ensures the obtaining of an exact solution in polynomial time. Then, the conclusion of the optimal trajectory can not be executed in an easy manner. To solve this problem, a particular family of algorithms called the heuristics, providing the obtaining of almost optimal solutions, is proposed [9]. For the KAT-mobility protocol, the local search algorithm 2-Opt, based on the modification of a current solution to TSP by heuristics, is implemented. Fortunately, for a WSN, sensors can communicate together, and the mobile sink does not need to visit all the sensors. Therefore, it is only needed to optimize the routes among group centers. The KAT-mobility approach assumes that the collector has a priori knowledge of the positions of its member sensors [11]. It is possible that it loses communications with its blocked or failed members. In this case, it can stay at the center of its cluster to discover broken sensors and its trajectory can then be recalculated as soon as it reaches the access point [11]. Accordingly, a new algorithm was proposed in [9], with the aim of providing a better and practical solution by improving the route optimization of the sink in the presence of faults and breakdowns. In order to ameliorate this algorithm performance and the conventional KAT-mobility algorithm, our purpose in this paper is to propose a priori strategy that ensures transmission reliability, achieves better energy efficiency and guarantees the obtaining of a solution in real time once the problem is disturbed by the failures of some sensors.

V. NKAT-MOBILITY PROTOCOL

We propose through this paper to perform a transmission of information in real time and in an automated manner via a WSN. Our proposal consists in presenting the New K-means and Traveling Salesman Problem based mobility (NKAT-mobility) protocol, based on K-means algorithm and the routing path optimization by using the Tabu method. The aim of this paper is to present this a priori strategy which ensures the reliability of data transmission in the presence of faults, so we model the wireless network by a graph G (V, E), where V is the set of group centers and E represents the set of edges reflecting the possible communications between these points. We denote by k the number of vertices in the graph, which is considering here as the problem size. In wireless sensor networks, one or several sensors may not function correctly, so the number of functional vertices varies between 0 and k. We are working in a situation where the vertices do not exist in a deterministic manner in this graph, but they are present in a probabilistic way. In other words, a probability of presence p_j will be associated with each vertex m_j (j = 1, 2 ... k). Moreover, on any instance of the problem, we avoid the total flaw of the system by changing the graph structure and transforming it into a subgraph, according to a modification strategy that will be specified in advance.

For the KAT-mobility protocol, the local search algorithm 2-Opt, based on the modification of a current solution to TSP by heuristics, is implemented. To ameliorate this work in the deterministic case, which is a particular case of this probabilistic problem, i.e., when p_j (j = 1, 2 ... k) = 1, we propose the improvement of the route optimization by implementing the Tabu algorithm instead of the used method 2-Opt. This algorithm depends on an initial solution, the neighborhood and a Tabu list. It is possible that the mobile sink lose communications with its failed or broken member nodes. In this case, it can stay at the center of its group to discover failed nodes and its trajectory can be recalculated as soon as it reaches the access point [11]. Updating this route is preceded by a modification of the centroids locations in the WSN. In other words, a new clustering procedure through the subset of operational sensors is performed. Consequently, we propose that only some centers among all the vertices will really necessitate a visit according to their probabilities of appearing (remaining functional) and we assume that when all a cluster members are failed, its center will be regarded as absent. We specify then a modification strategy μ, which removes absent centers from the initial a priori trajectory. This is a problem of finding a priori trajectory that minimizes the functional of covered distances. Our aim consists in obtaining a tour among the initial vertices such that G is transformed into the subgraph G'= G [V'], where V' ⊆ V is the set of present group centers and the new path through its vertices will be in the same order as that created by the a priori tour. Given the set of cluster centers, the probability law ℙ, the set of all the subsets of V, i.e., each instance V' ⊆ V has a probability of presence ℙ (V'). For a given trajectory R through the vertices defined on V, the modification strategy μ used for generating a realizable solution through V', consists in gumming or eliminating those who are absent from the a priori trajectory. Let $L_{(R,\mu)}$ be the random variable defined on 2^V, which for a trajectory R and for all V' ⊆ V, associates the length $L_{(R,\mu)}$ (V') through V', induced of the trajectory R by the modification method. Therefore, the route optimization of the mobile sink to visit once and just once every node is equivalent to find the trajectory that minimizes the functional of $L_{(R,\mu)}$ [21]. To solve our cited problem, this a priori strategy is modeled with the aim to maintain network functionality and to avoid the total system flaw despite of the existence of faults in a subset of its elementary components by minimizing the objective function of this problem. So, we propose the implementation of the Tabu algorithm for optimizing the routing path of the sink, as a second procedure of NKAT-mobility protocol, in the deterministic case as well as in the probabilistic case (in presence of breakdowns, i.e., when p_j (j = 1, 2 ... k) ≠ 1). The functional of covered distances depending on R and μ, can be expressed by:

$$\min_R \left(\mathbb{E}\left(L_{(R,\mu)}\right) = \sum_{V' \subseteq V} \mathbb{P}\left(V'\right)\ L_{(R,\mu)}\left(V'\right) \right) \quad (8)$$

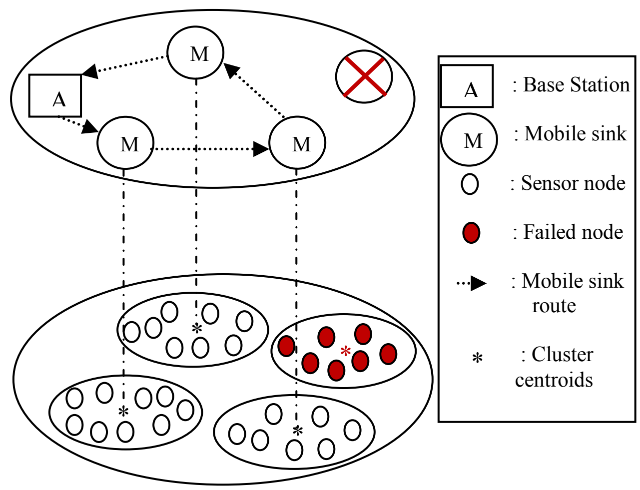


Figure 6. NKAT-mobility protocol

NKAT-mobility protocol allows us to have a real time solution, which is valid in all situations. This a priori strategy is used to achieve better performance than other existing methods and the new route obtained through the set of functional centroids is illustrated in Figure 6.

VI. Simulation Results and Analysis

In this section, simulation results are presented and analyzed. In fact, simulation is a very important step that enables us to analyze a WSN performance [39]. In fact, it is necessary to reduce the various possible conception errors by performing a validation step. To achieve our simulations and to study our approach (NKAT-mobility) performance, we have used the software Java. We have simulated the KAT-mobility protocol, the existed strategy proposed by Bellalouna et al. [9] and the NKAT-mobility protocol. In fact, we have implemented the K-means algorithm that represents the first module of each protocol and thereafter the second module of each one, which differs from one method to the other.

The WSN that we have simulated comprises 200 sensor nodes, which are dispersed inside the simulation area. They are deployed according to a random distribution in order to test our proposed strategy performance under conditions that are more similar to those real. We consider the metric tour cost (expressed in meter) for estimating the tour lengths.

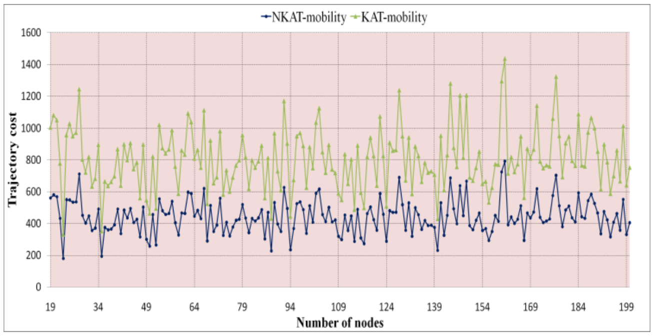


Figure 7. Variation of the trajectory cost according to the number of nodes in deterministic case

Through Figure 7 that represents the trajectory cost depending on the number of nodes, varying from 19 to 200, we can see that the tour length obtained by NKAT-mobility protocol for any used distribution is lower than that obtained by KAT-mobility protocol. Then, we can conclude that the new method based on K-means clustering and the approximate solution for TSP by using the local search method Tabu achieves better performance than the conventional method based on K-means clustering and the approximate solution for TSP by employing the simple local search algorithm 2-Opt, in deterministic case. In other words, NKAT-mobility protocol ensures the minimization of the trajectory length for any employed distribution, when the probability of presence p_j, associated with each vertex m_j ($j = 1, 2 ... k$), is equal to 1 (all the nodes are operational).

The nodes number may vary from one day to the other because of the possible failures and attacks, so the NKAT-mobility protocol was proposed to deal with this problem (probabilistic case). To confirm our a priori strategy performance, we have executed the simulations of the three mentioned algorithms: KAT-mobility, NKAT-mobility and the existed algorithm proposed by Bellalouna et al. [9].

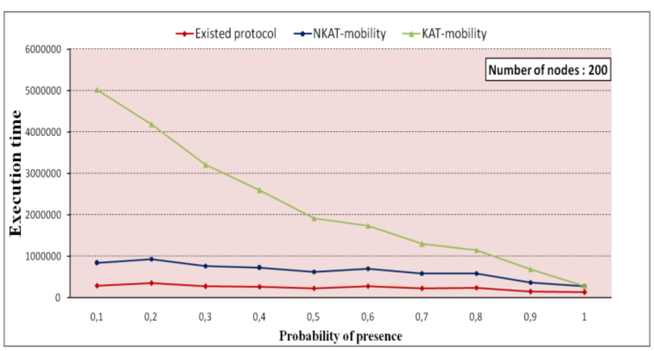


Figure 8. Variation of the execution time according to the probability of presence in probabilistic case

It is shown from the simulation results of Figure 8, representing the time execution, which is expressed in nanoseconds, depending on the presence probability of cluster centroids, that the NKAT-mobility protocol provides less execution time than the KAT-mobility protocol. Our proposed probabilistic model is much more realistic, once the problem is disturbed by the failures of some nodes (i.e., the probability of remaining functional is less than 1). It is clear from the simulation results that the reduction in terms of time, compared with the conventional method, is very important. However, we can notice a small gap between our protocol and the existed protocol. Therefore, it was interesting to compare these two methods by evaluating the functional of covered distances.

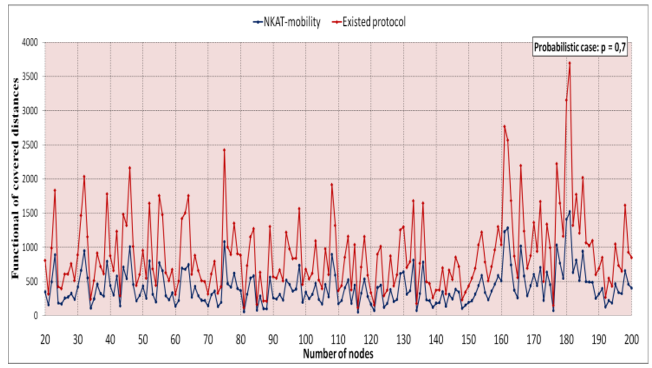


Figure 9. Performance comparison with failed nodes (30% absent)

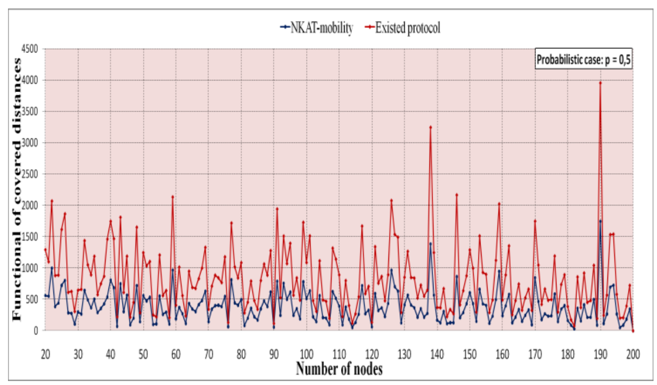


Figure 10. Performance comparison with failed nodes (50% absent)

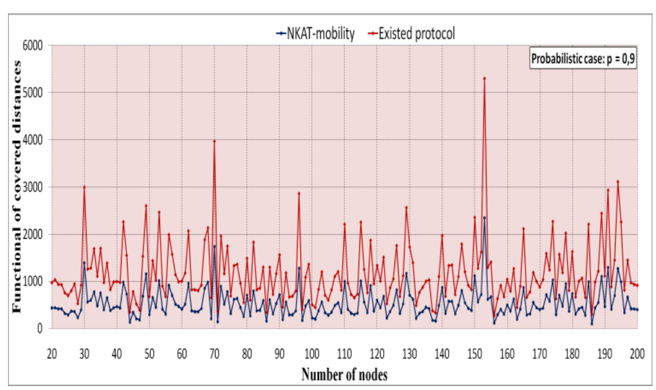


Figure 11. Performance comparison with failed nodes (10% absent)

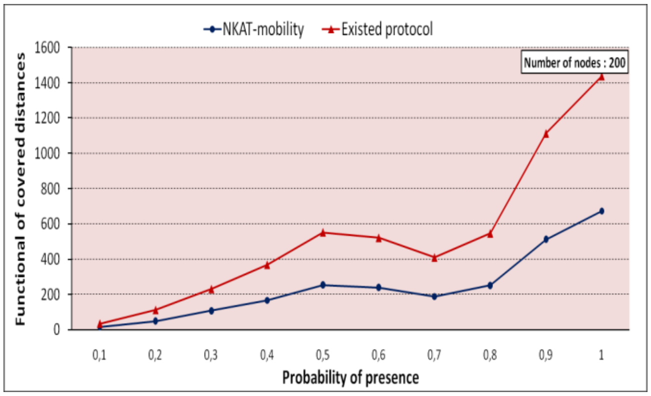


Figure 12. Variation of the functional according to the probability of presence in probabilistic case

Through Figures 9, 10 and 11, we can see that the functional of covered distances obtained by our method (NKAT-mobility) is lower than that of the existed strategy proposed by Bellalouna et al. [9], for any used distribution and for any tested values of p_j. In fact, these figures representing the functional variation, with values of n ranging from 20 to 200 and with different values of the presence probability (the percentages of centroids absence are 10%, 30%, 50%), shows that NKAT-mobility achieves better performance than the existed protocol and guarantees the functional minimization. This idea is also confirmed in Figure 12 that represents the functional variation when p_j varies and when n is fixed at 200. It is clear from our simulations that the improvement provided by our new method is effective and that our algorithm is more realistic than the fault tolerant KAT-mobility protocol, valid in all situations and better than the existed protocol in terms of solution quality (optimized routing path).

VII. Conclusion and Future Works

In this paper, a new strategy of data collection in a WSN has been proposed. The novelty of this scheme is that the combinatorial optimization provides applicable approaches in the context of wireless sensor networks. In other words, the WSN can be modeled as a probabilistic combinatorial optimization problem, particularly in presence of breakdowns. Then, we have presented the fault tolerant NKAT- mobility protocol that assures finding a real time solution to each new problem instance, once it is disturbed by the absence of some nodes. This a priori strategy aims to provide better practical solution compared with two conventional methods based on K-means clustering and the approximate solution for TSP, in probabilistic case. Simulation results demonstrated that the performance differences between our probabilistic model and the two methods are significant. We can conclude that NKAT-mobility ensures the improvement of the collector route optimization. To accomplish this work, we aim to use a new strategy by implementing various algorithms employed for probabilistic combinatorial optimization problems resolution, such as approximate methods and this will be so interesting when the risk of breakdowns becomes considerable and the number of cluster centroids will be reduced.

References

[1] S. Clement, “Some contributions in wireless sensors networks: Localization and Routing,” Ph.D thesis, Avignon University, France, 2008.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” Computer Networks, vol. 38, issue 4, 2002, pp. 393-422, doi: 10.1016/S1389-1286(01)00302-4.

[3] H. Bousetta, “Multiphysics modeling and global simulation of autonomous systems on chip,” Ph.D thesis, National Polytechnic Institute of Grenoble, France, 2010.

[4] K. Bouchakour, “Hierarchical routing in wireless sensor networks: KhLCH protocol (K-hop Layered Clustering Hierarchy),” Ph.D thesis, National High School of Computer Sciences, Algeria, 2012.

[5] M. Hempstead, M. J. Lyons, D. Brooks, and G. Wei, “Survey of hardware systems for wireless sensor networks,” Low Power Electronics, vol. 4, issue 1, 2008, pp. 11-20, doi: 10.1166/jolpe.2008.156. 2008.

[6] M. Bellalouna, “Probabilistic combinatorial optimization problems,” Ph.D thesis, National School of Bridges and Roads, Paris, 1993.

[7] M. Bellalouna and A. Ghabri, “A priori methods for fault tolerance in wireless sensor networks,” World Congress on Computer and Information Technologies (WCCIT), IEEE, Sousse, 2013, pp. 1-6, doi: 10.1109/WCCIT.2013.6618654.

[8] C. Karlof and D. Wagner, “Secure routing in wireless sensor networks: attacks and countermeasures,” Ad Hoc Networks, vol. 1, 2003, pp. 293-315, doi: 10.1016/S1570-8705(03)00008-8.

[9] M. Bellalouna, A. Ghabri, and W. Khaznaji, “The K-means and TSP Based Mobility Protocol Modeling as a Probabilistic Combinatorial Optimization Problem,” in proceedings of the Ninth International Conference on Systems and Networks Communications (ICSNC), IARIA, Nice, France, October 2014, pp. 48-53.

[10] D. J. Bertsimas, “Probabilistic combinatorial optimization problems,” Ph.D thesis, MIT, Cambridge, Mass, 1988.

[11] H. Nakayama, N. Ansari, A. Jamalipour, and N. Kato, “Fault-resilient sensing in wireless sensor networks,” Computer communications, vol. 30, issue 11, September 2007, pp. 2375-2384, doi: 10.1016/j.comcom.2007.04.023. 2007.

[12] P. Jaillet, “Analysis of Probabilistic Combinatorial Optimization Problems in Euclidean Spaces,” Mathematics of Operations Research, vol. 18, issue 1, February 1993, pp. 51-70, doi: 10.1287/moor.18.1.51.

[13] D. Bertsimas, P. Jaillet, and A. Odoni, “A priori Optimization,” Operations Research, vol. 38, issue 6, December 1990, pp. 1019-1033, doi: 10.1287/opre.38.6.1019.

[14] A. Henchiri, M. Bellalouna, and W. Khaznaji, “Probabilistic traveling salesman problem: a survey,” in proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS), Warsaw, September 2014, pp. 55-60, doi: 10.15439/2014F381.

[15] P. Jaillet, “Probabilistic Traveling Salesman Problems,” Ph.D thesis, MIT, Cambridge, Mass, USA, 1985.

[16] P. Jaillet, “Analysis of the probabilistic traveling salesman problem in the plane,” Ricerca Operativa, vol. 36, January 1986, pp. 57-68.

[17] C. Murat and V. Paschos, “The probabilistic longest path problem,” Networks, vol. 33, issue 3, 1999, pp. 207-219, doi: 10.1002/(SICI)1097-0037(199905)33:3<207::AID-NET7>3.0.CO;2-7.

[18] C. Murat and V. Paschos, "The probabilistic minimum vertex covering problem," International Transactions in Operational Research, vol. 9, issue 1, January 2002, pp. 19-32, doi: 10.1111/1475-3995.00338.

[19] Z. Bouyahia, M. Bellalouna, P. Jaillet, and K. Ghedira, "A Priori Parallel Machines Scheduling," Computers & Industrial Engineering, vol. 58, issue 3, April 2010, pp. 488-500, doi: 10.1016/j.cie.2009.11.009.

[20] M. Bellalouna and L. Horchani, "The Two-Dimensional Probabilistic Bin-Packing Problem: An average case analysis," International Journal of mathematics and computer in simulation, vol. 2, issue 1, March 2008, pp. 42-46.

[21] P. Jaillet, "A priori solution of a Travelling Salesman Problem in which a random subset of the customers are visited," Operations research, vol. 36, issue 6, Maryland, US, 1988, pp. 929-936, doi: 10.1287/opre.36.6.929.

[22] M. Bellalouna, C. Murat, and V. Paschos, "Probabilistic combinatorial optimization problems on graphs: A new domain in operational research," European Journal of Operational Research, vol. 87, issue 3, December 1995, pp. 693-706, doi: 10.1016/0377-2217(95)00240-5.

[23] L. Forlizzi, J. Hromkovic, G. Proietti, and S. Seibert, "On the stability of approximation for hamiltonian path problem," in proceedings of the Conference on Current Trends in Theory and Practice of Computer Science, Berlin, 2005, pp. 147-156, doi: 10.1007/978-3-540-30577-4_18.

[24] D. J. Bertsimas and L. H. Howell, "Further results on the probabilistic traveling salesman problem," European Journal of Operational Research, vol. 65, issue 1, February 1993, pp. 68-95, doi: 10.1016/0377-2217(93)90145-D.

[25] D. J. Bertsimas, "A Vehicule Routing Problem with Stochastic Demand," Operational Research, vol. 40, issue 3, June 1992, pp. 574-585, doi: 10.1287/opre.40.3.574.

[26] M. Gendreau, G. Laporte, and R. Séguin, "A Tabu Search Heuristic for the Vehicle Routing Problem with Stochastic Demands and Customers," Operations Research, vol. 44, issue 3, May 1996, pp. 469-477, doi: 10.1287/opre.44.3.469.

[27] S. Rosenow, "Comparison of an Exact Branch-and-Bound and an Approximative Evolutionary Algorithm for the Probabilistic Traveling Salesman Problem, Operations Research, August 1998, pp. 168-174, doi: 10.1007/978-3-642-58409-1_16.

[28] D. S. Johnson and L. A. McGeoch, "The Traveling Salesman Problem: A Case Study in Local Optimization," John Wiley & Sons, New York, NY, USA, 1997, pp. 215-310.

[29] M. Bellalouna, V. Paschos, and W. Khaznaji, "Well solved cases of probabilistic traveling salesman problem," 42nd days of statistics, INRIA, Marseille, France, June 2010, pp. 1-7.

[30] P. Jaillet, "On some probabilistic combinatorial optimization problems defined on graphs," Flow Control of Congested Network, vol. 38, 1987, pp. 255-267, doi: 978-3-642-86726-2_16.

[31] L. Bianchi, J. Knowles, and N. Bowler, "Local search for the probabilistic traveling salesman problem: Correction to the 2-p-opt and 1- shift algorithms," European Journal of Operational Research, vol. 162, issue 1, February 2005, pp. 206-219, doi: 10.1016/j.ejor.2003.10.016.

[32] L. Bianchi and A. M. Campbell, "Extension of the 2-p-opt and 1-shift algorithms to the heterogeneous probabilistic traveling salesman problem," European Journal of Operational Research, vol. 176, issue 1, January 2007, pp. 131-144, doi: 10.1016/j.ejor.2005.05.027.

[33] F. Glover, "Future paths for integer programming and links to artificial intelligence," Computer & Operations Research, vol. 13, issue 5, 1986, pp. 533-549, doi: 10.1016/0305-0548(86)90048-1.

[34] H. Nakayama, N. Ansari, A. Jamalipour, Y. Nemoto, and N. Kato, "On data gathering and security in wireless sensor networks," in proceedings of the IEEE Sarnoff symposium, IEEE, Nassau Inn, Princeton, NJ, 2007, pp. 1-7, doi: 10.1109/SARNOF.2007.4567337.

[35] J. P. Jafrin and P. A. C. Angelin, "A comparative study of data gathering algorithms for a mobile sink in wireless sensor network," International journal of advanced research in computer engineering and technology, vol. 1, issue 9, November 2012, pp. 250-255.

[36] Z. Guellil and L. Zaoui, "Proposition of a solution to the initialization problem K-means Case," in proceedings of the second International Conference on Informatics and Applications (CIIA), Saida, Algeria, 2009, pp. 1-9.

[37] P. Mathur, S. Saxena, and M. Bhardwaj, "Node clustering using K Means Clustering in Wireless Sensor Networking," in proceedings of the second National Conference in Intelligent Computing Communication (EICC), Department of Computer Science and Engineering, Indian Institute of Technology, Uttar Pradesh, India, 2013, pp. 1-6.

[38] T. Fukabori, H. Nakayama, H. Nishiyama, N. Ansari, and N. Kato, "An efficient data aggregation scheme using degree of dependence on clusters in WSNs," in proceedings of the IEEE International Conference on Communications, May 2010, pp. 23-27, doi: 10.1109/ICC.2010.5502285.

[39] A. Makhoul, "Sensor Networks: localization, coverage and data fusion," Ph.D thesis, University of Franche-Comte, France, 2008.

New Geometric Geolocalization Method for Tracking Patients in Medical Environments

Afef Ghabri[1], Monia Bellalouna[2], Leila Horchani[3]
[1]Laboratory RIADI-GDL, [23]Laboratory CRISTAL POLE GRIFT
National School of Computer Sciences (ENSI)
Manouba, Tunisia
e-mail: [1]afef2108@hotmail.com, [2]monia.bellalouna@ensi.rnu.tn, [3]leila.horchani@gmail.com

***Abstract*— The geolocalization is a technique which assures the determination of the geographical positions of individuals through their mobile equipments or elements equipped with a specific device. It has received research interest due to the success of the emerging wireless sensor network technology. To achieve some applications, such as environmental monitoring, military surveillance and disaster relief, node localization is fundamental and essential. Moreover, the availability of location information can enable a myriad of applications such as, intrusion detection, inventory management, road traffic monitoring, reconnaissance, health monitoring and surveillance, in order to insure an intervention in the suitable time and place. There are different methods of geolocalization able of resolving this problem. All these approaches are characterized by their own disadvantages and advantages, making them adapted to various applications. So, we always try to improve them by proposing a new approach with the aim of having a better precision and a more accurate positioning. This paper presents a new geometric method which can improve the poor locating performance of the typical Distance Vector-Hop (DV-Hop) algorithm, without increasing the hardware cost for sensors. We propose this approach to ensure the monitoring and the positioning of patients in medical environments for indoor use and to achieve a good compromise between positioning accuracy, computation complexity and energy cost. The influences of anchors on the algorithm and simulation results are also explored in this paper.**

Keywords-Geolocalization; wireless sensor network; precision; Distance Vector–Hop; medical environments.

I. INTRODUCTION

Wireless Sensor Networks (WSN) represent a very interesting domain of development that has become a new research focus in communication and computer fields [1]. They are tremendously being used in different environments, in a random way, to perform various monitoring tasks such as rescue, search, target tracking, disaster relief and a number of tasks in smart environments [2]. A wireless network is composed of plenty of nodes that are deployed in the monitoring field, and that have perception, processing and communication ability [3]. Each node is capable of sensing the environment and then communicating the measured data to its neighbours, and eventually to the external users. Many applications require geographical information to work effectively. Indeed, we cannot make decisions without knowing the locations of some persons or certain objects because there are information that depends on the positions of the users like the forecasts, the meteorological observations and also the road traffic [4]. Localization capability is necessary in most WSN applications, particularly for tracking patients in medical environments. In fact, the determination of a person location means placing it in a defined environment or finding its place in a space with Cartesian coordinates. So, one of the basic challenges in the WSN is node geolocalization, which is a particular case of localization [2][5]. It is the positioning of a person or an object using geographic locations. Thus, we can distinguish between so many approaches to assure the implementation of this process. These approaches commonly aim at positioning target nodes as accurately as possible. A localization method is a set of technologies implemented and used to perform the positioning task [4]. Its performance depends on some factors, such as the system accuracy, the costs of communication and the density of nodes.

The purpose of this paper is to explain the principle and the interest of the specific localization case through geographic coordinates. The rest of the paper will be organized as follows: In the second section, we will present an overview of the geolocalization problem. A presentation of the problem in medical environments will be also cited in Section 3. Section 4 describes the basic DV-Hop algorithm. In the next section, we discuss a new geometric positioning approach based on DV-Hop algorithm. Section 6 presents the simulation results and analysis. Finally, we will conclude the paper in Section 7.

II. WIRELESS SENSOR NETWORK GEOLOCALIZATION

Many applications use a random deployment of a large number of sensors. So, the localization phase is required to realize network operation and also to exploit collected information. We must then locate all the nodes with the best possible accuracy. Just like management issues of energy consumption, the problem of localization and specifically of geolocalization seems challenging and research works should focus on the development of effective positioning methods. In fact, the geolocalization in the WSN is considered as a typical case of localization. It presents a process that ensures the positioning of an object (information, a person, etc.) on a map or a plan through its

 ISBN: 978-1-61208-421-3

geographic coordinates. It is also used in order to position mobile nodes. The realization of this operation is done using a terminal capable of being located by using a Global Positioning System (GPS) receiver or through other techniques and ensuring the publication of its geographical coordinates in a differed way or in real time [5]. Saved locations can be stored inside the terminal and extracted later, or transmitted in real time towards a software platform that provides the geolocalization. Sending in real time requires a terminal equipped with a telecommunication tool which its type can be radio or satellite, allowing it to transmit the positions during regular intervals. Thus, the terminal location can be visualized in a map by a geolocation platform which is often accessible via Internet.

Geolocalization is needed today for the development. It is applicable in several levels. It allows tracking vehicles carrying people and goods and ensures their localization in real time and also the detection of problems. In fact, these are sensors with varied forms, making it possible to provide information according to geographical positions of the users. The geolocalization offers many services in different fields. For example, security services can find reinforcement in some positioning techniques. Thus, everyone is currently agreed on its important value.

Before its establishment, the WSN deployment requires a simulation phase of the used protocol to ensure a good performance. There are so many solutions for WSN positioning that can be classified according to various criteria. These methods include Amorphous [6][7], Centroid [8], DV-Hop [9] and Approximate Point In Triangulation (APIT) [10]. These techniques are simple because they respect the material and energy constraints of sensor nodes and they also provide a low cost. However, their fundamental problem is the low accuracy. Thus, the main objective of this paper is to propose a geolocalization technique for tracking patients in health care settings, ensuring high accuracy without increasing the operational cost of the network.

III. Geolocalization in Medical Environments

In the sociomedical institutions, clinics and hospitals, ensuring the security and the protection of patients, as well as safeguarding mobile equipments, has become an increasing priority. Currently, monitoring and supervision of patients is done manually in hospital by medical staff and in a regular way [11]. This process of collecting information can be problematic. In fact, the collection is not done in real time (so many times per day) and hence the staff mobilization becomes necessary during a considerable interval of time. Thus, one often tries to solve these problems while benefiting from technological advances that provide advantageous techniques of care and assistance for the patients. The WSN is used in the medical field as it allows patient monitoring at a distance and the control of human physiological data, such as heart rate and glucose level [12]. Indeed, we can attach to the patient small sensor nodes able to carry out specific tasks, such as detecting heartbeats and the control of blood pressure. In addition, video micro-sensors can be implanted under the skin in order to receive images of a body part in real time without surgery. Therefore, it will be possible to supervise the recovery of muscles or the progression of diseases. Several medical problems, such as cancer, can be then controlled. Wireless sensors permit also the detection of older people behavior, so one can intervene quickly in case of need [13]. Staff can be informed of the transmitter location and also the patient name, and as soon as a disorientated patient or a sick person suffering from senile dementia or psychic disorders leaves a protected area, staff will be warned using alert messages that can be sent via a wireless telephone, an alarm system, a pager or the call nursing system.

The sensors are used not only for the wireless monitoring of the patients, but also for the monitoring and the control of the doctors within the hospital. In addition, they can be attached to drugs and medicines to be administered, in order to decrease the probability of giving bad treatments to patients and minimizing the possibilities of side effects caused by the use of inappropriate drugs. The sensor network is effective in the medical field, as the micro-sensors provide a movement ease for the controlled and monitored subjects and also a faster identification of some symptoms to doctors. We propose through this paper to perform a collection of information in real time and in an automated manner via a WSN. Our proposal consists of presenting a global method for tracking and positioning patients in health care settings, grouping several technologies, for an indoor employment.

In the field of WSN geolocalization, we always try to guarantee a better precision with a low energy consumption and lower cost. We try then to benefit from the advantages of the various existed localization methods and we propose a new algorithm which solves the problems of the constraints satisfaction. This problem can be resolved by employing an algorithm which consists in making the combination of different techniques. Indeed, it exploits the good properties of the various methods by applying them to the problems that can be solved efficiently [14][15]. As a typical range-free algorithm [4], DV-Hop algorithm has the advantage to localize normal nodes that have less than three neighbour anchors. Many works in the literature propose schemes and improved DV-Hop methods for many purposes. Since each was developed to fulfill a different goal, they vary widely in many parameters including accuracy, cost, configurability, security, and reliability [3].

IV. DV-Hop Positioning Algorithm

Niculescu and Nath [16] have proposed the DV-Hop algorithm, which is one of the typical representatives of range-free localization algorithm. Its basic idea is that the distance between the unknown nodes and the reference nodes is expressed [17]. The algorithm implementation is comprised of three steps. First, it employs a classical distance vector exchange, so that all nodes in the network get distances, in hops, to the landmarks. Then, it estimates an average size for one hop, which is deployed as a correction to the entire network. Finally, unknown nodes compute their location by trilateration [18]. The complexity of this algorithm is higher than that of some approaches, such as Centroid [19] and Convex Position Estimation (CPE) [20], designed for normal nodes which have at least three neighbour anchors. The Centroid algorithm has a low computation cost, and does not increase the network traffic. It can also get relatively good accuracy, when the distribution of anchors is regular. However, when the distribution of anchors is not even, the estimated position derived from the Centroid algorithm will be inaccurate [21]. CPE has slightly higher localization accuracy than Centroid [3].

DV-Hop is a suitable solution for normal nodes having less than three neighbour anchors. In the first implementation step, each anchor node broadcasts an anchor to be flooded throughout the network containing the anchors location with a hop count value initialized to one. Each receiving sensor maintains the minimum hop count value. Anchors with higher hop count values are defined as the useless and ignored information. Then, the rest of anchors are flooded outward with hop count values incremented at every intermediate hop. Through this process, all nodes in the network get the minimal hop count to every anchor node. In the second step, once an anchor gets hop count value to other anchors, it estimates an average size of one hop, which is then flooded to the entire network. After receiving the hop size, the unknown nodes multiply the hop size by the hop count value to derive the estimated physical distance from the anchor. Each anchor node broadcasts its hop size to the network using controlled flooding. Unknown nodes receive the hop size information and save the first one. Meanwhile, they transmit the hop size to their neighbour nodes. This scheme ensures that most nodes receive the hop size from the anchor that has the least hops between them. In the last step, when unknown nodes get three or more distance information from anchors, the trilateral measuring method or the maximum likelihood estimation approach is used to calculate their locations [3]. The average hop size is estimated by the anchor node i using the following formula [22]:

$$HopSize_i = \frac{\sum_{j \neq i} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{j \neq i} h_{ij}} \qquad (1)$$

where (x_i, y_i) and (x_j, y_j) are coordinates of anchors i and j, respectively, and h_{ij} is the hop count between the anchor nodes i and j. It is assumed that (x_u, y_u) are the location of the unknown node u, (x_i, y_i) are the known location of the ith anchor node and d_{ui} is the distance between them. Then, the coordinates of u are computed by the following formula [17]:

$$F = (C^T C)^{-1} C^T D \qquad (2)$$

where
$$F = \begin{bmatrix} x_u \\ y_u \end{bmatrix} \qquad (3)$$

$$C = -2 \times \begin{bmatrix} x_1 - x_n & y_1 - y_n \\ \vdots & \vdots \\ x_{n-1} - x_n & y_{n-1} - y_n \end{bmatrix} \qquad (4)$$

and
$$D = \begin{bmatrix} d_{u1}^2 - d_{un}^2 - x_1^2 + x_n^2 - y_1^2 + y_n^2 \\ d_{u2}^2 - d_{un}^2 - x_2^2 + x_n^2 - y_2^2 + y_n^2 \\ \vdots \\ d_{u(n-1)}^2 - d_{un}^2 - x_{n-1}^2 + x_n^2 - y_{n-1}^2 + y_n^2 \end{bmatrix} \qquad (5)$$

Finally, the refinement process of locations is carried out. In fact, after estimating their positions, the sensors ensure their dissemination to their neighbors. Thus, based on this information and according to the different neighborhood relations, these sensors determine again their positions which are close to their actual locations. The sensors fix then their estimated positions after a definite number of iterations. The precision of the DV-Hop algorithm is low because the average distance of one hop represents an approximate estimation and it is difficult to know the exact values of the distances separating the nodes [23]. Indeed, this algorithm exactitude varies according to the precision of these distances. The second weakness is related to the type of nodes distribution; because the more regular it is, the more results are accurate, i.e., it is not applicable for the random distributions. However, the advantage of this method is that it doesn't require additional hardware for the distances estimate and the determination of the unknown locations. This significantly reduces the economic and energetic localization cost and also it avoids the constraint imposed by Centroid and CPE (the requirement of a minimum of three neighbour anchors). Moreover, it employs a very simple algorithm that does not require too many computing and radio communication and processing operations, i.e., not much energy consumption.

V. IMPROVED DV-HOP GEOLOCALIZATION SCHEME

We propose through this paper to perform a collection of information in real time and in an automated manner via a WSN. Our proposal consists of presenting a global method for tracking and positioning patients in health care settings, grouping several technologies, for an indoor employment. In fact, one or several fixed nodes will form a backbone which is responsible for relaying the data collected by sensors that are placed on mobile patients. Fixed sensors must be organized in an efficient and robust virtual structure, whereas the other mobile nodes require being able to be adaptable to various displacements in order to broadcast the information in a proper manner. It is essential to use mobile anchors that will communicate with the other nodes to simplify the coordinates attribution to the sensors attached to patients. Generally, the existence of several reference nodes is compulsory to have a good accuracy and certainty in the determination of the positions but practically this is difficult because of the rising cost of these anchors. In the wireless

 ISBN: 978-1-61208-421-3

sensors network, anchors are necessary for the localization of nodes in a global coordinates system. These are ordinary nodes that know their coordinates a priori. These anchors, which play a very important role, can be affixed or carried by people working at the hospital (doctors, nurses, etc.) and they also can be inserted in medical equipments. Figure 1 shows this principle.

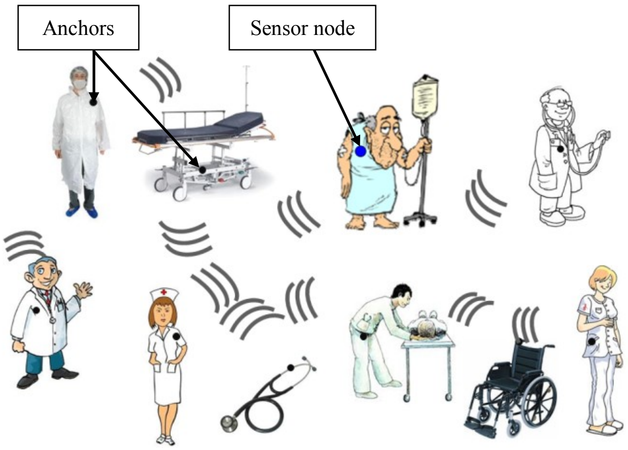


Figure 1. Positioning method in medical environments

Since their positions change over time, it is necessary to update them. To do this, we propose to employ Radio Frequency Identification (RFID) technology. In fact, it is based on remote research, extraction and storage of data, which requires the presence of an RFID reader, a tag and a database. It designates the recognition of objects and even people by radiofrequency transmission [24]. An RFID system includes a reader (interrogator) that transmits a signal to a radio tag or more placed in its reading field depending on a given frequency. This is a transceiver which provides the communication with the corresponding tag by electromagnetic waves. This latter is designed to receive a radio signal and then to return immediately another response signal containing relevant information. To perform data exchange, a dialogue is established between these two components using a communication protocol. Tags are small objects that can be incorporated into products or implanted in living beings like animals or the human body. They include an antenna linked to an electronic chip and this antenna allows them responding to requests sent from the reader. RFID system is illustrated in Figure 2.

Figure 2. RFID system (reader+tag)

Indeed, a set of RFID readers that are equipped with several types of antennas are placed in order to cover the entire desired area. This zone is then subdivided into cells with varied surfaces according to the power of the deployed readers and their number. If a person equipped with RFID tag is located in these zones, it will be possible that the system calculates its position based on the number of readers detecting the tag, and it will be able to deduce the approximate location of the individual on the basis of the established cutting diagram [25]. However, this technique remains very approximate in real time and its exactitude will only specify the hallway or the part where the person exists. This geolocalization technique is especially employed in hospitals as RFID readers can be positioned in the building doors in order to indicate if a person, equipped with a tag, is going through them. Then, we will choose to use powerful readers and passive tags, because they are characterized by a low cost and a small size and their functional lifetime is important. Each tag will be associated with an anchor and each time it will be awake by a reader, a dialogue will be established and data will be exchanged. An update of this anchor position will be made and we will be able to use it in order to geolocate the patients. Figure 3 shows this principle.

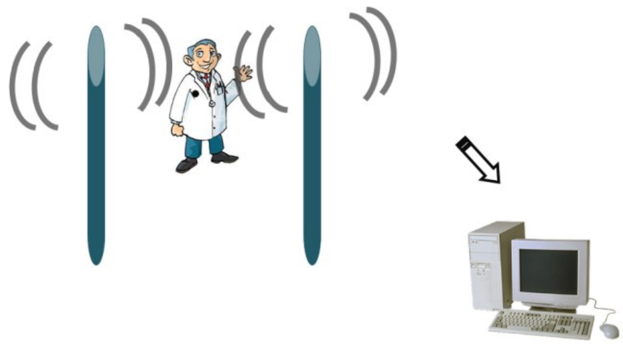


Figure 3. Data transmission procedure

In the field of WSN geolocalization, we always try to guarantee a better precision with a low energy consumption and lower cost. We try then to benefit from the advantages of the various existed positioning methods and we propose a new algorithm which solves the problems of the constraints satisfaction. This algorithm must find the most economical path and ensure a prolongation of the nodes lifetime [26]. Our approach consists in proposing an algorithm based on DV-Hop which is considered as a typical example of the category of methods that are at once distributed, anchor-based and range-free [4]. Indeed, the DV-Hop method is not precise enough [22]. Thus, the localization accuracy must be further improved. Then, we seek to guarantee the reduction of our algorithm cost and to ensure assessment of the network load caused by the protocol which is associated with this algorithm. Consequently, we propose a new algorithm in order to realize a satisfactory compromise between location accuracy and computation complexity. Our improved DV-Hop (Intersect DV-Hop) consists of four steps. The first two steps are the same as the DV-Hop method, while others are specific to our proposal.

In step one, all nodes in the network get the minimal hop number to every anchor. In step two, after obtaining the hop-size, anchor node broadcasts its hop-size to network as a correction. In fact, at this point, each normal node N_x knows its $hop_{i,Nx}$, which represents the minimal hop number from N_x to A_i. Each anchor A_i has also obtained $hop_{i,k}$, which is the minimal hop number to any other anchor A_k. So, A_i can calculate its average distance per hop dhp_i, and then broadcasts dhp_i through the network. After receiving dhp_i, the normal node N_x can use (6) to get $d_{i,Nx}$, which is the approximate distance between each anchor A_i and N_x.

$$d_{i,Nx} = hop_{i,Nx} \times dhp_i, \quad i = 1,2....m \qquad (6)$$

Our aim in the third step is to determine the closest three anchors to the normal node N_x by comparing the approximate distances obtained at the end of the previous step. The node is then located in the coverage area of cells of these anchors. Let $A_1(x, y)$, $A_2 (x_1, y_1)$ and $A_3(x_2, y_2)$ be these three beacons, knowing that $d_{1,Nx} < d_{2,Nx} < d_{3,Nx}$.

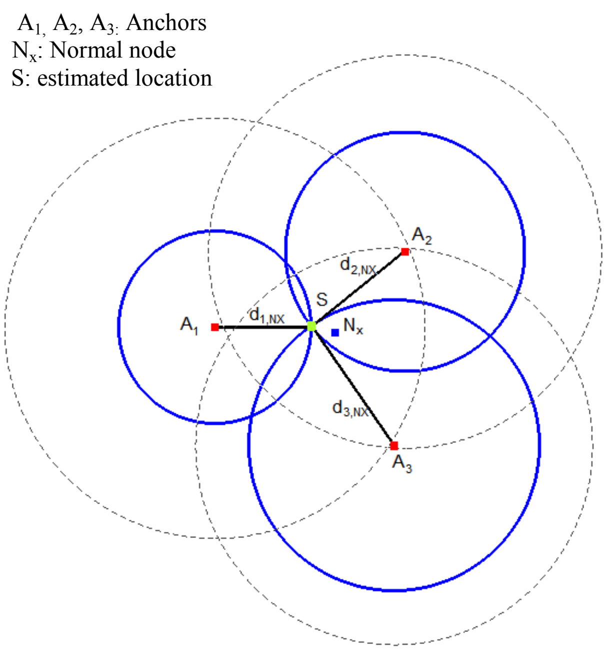


Figure 4. Principle of Intersect DV-Hop

We assume that the communication ranges of anchors are all the same as displayed by the circles of Figure 4. The normal node N_x locates in the overlapping communication region of A_1 A_2 A_3. We determine then at the last step the intersection of the three circles of Figure 4 that their centers are respectively A_1, A_2 and A_3 and their radius are respectively $d_{1,Nx}$,$d_{2,Nx}$ and $d_{3,Nx}$. Finally we calculate S, which is the final estimated position of N_x. If the coordinates of three anchors A_1 A_2 A_3 are (x, y), (x_1, y_1), and (x_2, y_2) and in the case where A_1 is considered as the origin of the coordinates system, the cross point S of the three circles can then be calculated as displayed by (7). Finally, S becomes N_x's estimated position which is the result of our proposed method.

$$\begin{cases} x_S = \dfrac{x_1^2 + d_{1,Nx}^2 - d_{2,Nx}^2}{2x_1} \\ y_S = \dfrac{x_2^2 + y_2^2 + d_{1,Nx}^2 - d_{3,Nx}^2 - 2x_S x_2}{2y_2} \end{cases} \qquad (7)$$

VI. SIMULATION RESULTS AND ANALYSIS

In this section, simulation results are presented and analyzed. The obtained results during the evaluation of the DV-Hop algorithm and the suggested improvement will be presented. To achieve our simulations and to study our approach performance, we have developed the Matlab software. The sensor network that we have simulated comprises 100 sensors (ordinary nodes and anchors) which are dispersed in a zone with the fixed size of 100 X 100 m^2 (Figure 5). They are deployed according to a random distribution in order to test the performances of the proposed algorithm under conditions more similar to those real. The random placement of the anchors allows measuring our approach exactitude in a more realistic context.

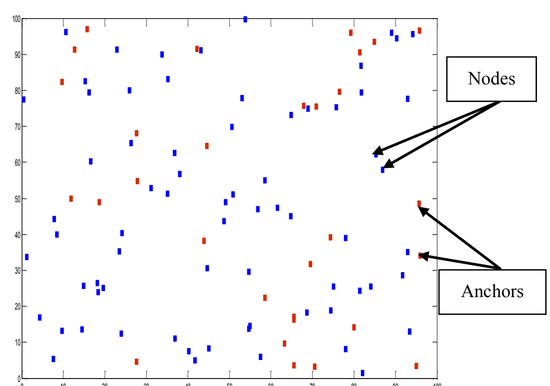


Figure 5. Random distribution of the sensors

The two most important parameters in our work are the error and the precision of localization. According to the various research studies already realized concerning the anchors influence on positioning, it was noted that the exactitude increases with the increase of anchors number. Therefore, we are interested in evaluating the variation of the precision and the error according to the number of employed anchors. The radio range of the sensors (R) indicates the communication range of nodes, which generally varies between 10 and 100 meters. Assuming the absence of the obstacles and the multiple paths, the antennas are considered to be ideal and the nodes keep the same radio range modulated by a circle that surrounds each sensor.

In our work, R is set to 50 meters. To quantify the accuracy of a geolocalization method, we consider the localization error it makes. Even for the best techniques, an inevitable error is associated with each one of them. In our simulations, it is regarded as the distance separating the real node position and that calculated using the geolocalization technique. The more this distance is considerable, the less the system is accurate. So, we consider the metric geolocation error (expressed in radio range percentage) for evaluating the performance of geolocalization accuracy. Geolocation error is defined as distance (estimated location, true location)/NB×100%. Here, distance (estimated location, true location) is the distance between a node's estimated position and its true position and NB is the number of sensors.

An anchor ratio is defined as the ratio of the number of anchors on the total number of nodes which is set to 100. In other words, it is the ratio of anchors among the network nodes. These ratios vary from 3% to 60%. We have simulated four algorithms: our Intersect DV-Hop algorithm and three existing algorithms, which are DV-Hop, Checkout DV-Hop [8] and Mid-Perpendicular algorithm [3].

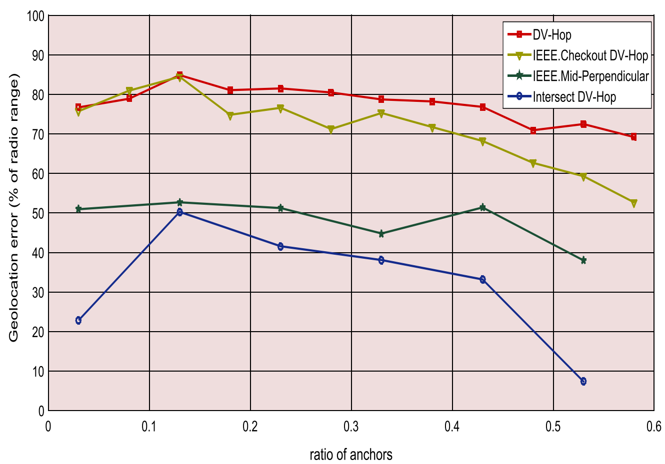


Figure 6. Geolocation error

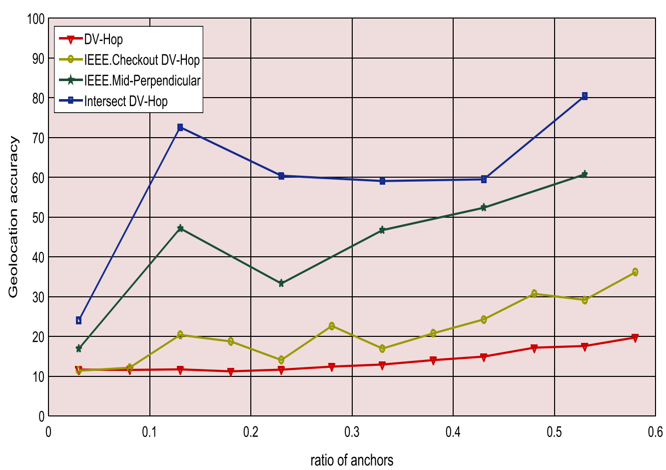


Figure 7. Geolocation accuracy

Through Figure 6, we can see that the average error of our algorithm is lower than that of other techniques. Then, we can conclude that the new method achieve better performance than Mid-Perpendicular, Checkout DV-Hop and DV-Hop. This idea is confirmed and also shown in Figure 7, which presents the positioning accuracy depending on ratio of anchor nodes for the four algorithms. We can notice that increasing the number of anchors guarantee the reduction of the localization error. For the same ratio of anchor nodes, position error is smaller when our improved method is applied in the same WSN environment than the DV-Hop algorithm and the two others. For example, with 10 anchor nodes (10%), Intersect DV-Hop has an average error of about 42% R, while the others are more than 51% R. When the number of used anchors increases to 50, the error decreases and becomes 17%. In this case, the geolocalization errors of the other algorithms are approximately more than 42%. It is clear from our simulations that the improvement provided by this new method is effective and that our algorithm is more accurate than the others, already mentioned. The Intersect DV-Hop performance exceeds the original DV-Hop location algorithm performance from our simulations results.

VII. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a new geolocalization algorithm for wireless sensor networks based on the DV-Hop algorithm. It was shown from the simulation results that our proposed algorithm can improve location accuracy than the original DV-Hop, Checkout DV-Hop and Mid-Perpendicular. The proposed method adopts the distance reliability information to improve the convergence rate and the location accuracy using a geometric approach and adaptive iterative process. We have to find a solution to fulfill this paper objective, which consists in proposing a method of tracking and monitoring patients in order to determine their geographic locations inside a medical environment. Our suggested approach is mainly based on the improvement of the DV-Hop performance with the use of RFID technique that seems very beneficial in this case. It was shown that our contribution is interesting and that it guarantees a good improvement in the geolocalization accuracy. To accomplish our work, we aim to test our solution on the practical level. In fact, this test will put it under more real conditions and consequently we can refine the searched locations. Finally, we also aim to propose a strategy that guarantees the obtaining of solutions in real time once the problem is disturbed by the failures of some nodes.

REFERENCES

[1] M. Bellalouna, A. Ghabri, and W. Khaznaji, "The K-means and TSP Based Mobility Protocol Modeling as a Probabilistic Combinatorial Optimization Problem," The Ninth International Conference on Systems and Networks Communications (ICSNC), Nice, France, 12 October 2014, pp. 48-53.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Computer Networks, vol. 38, issue 4, 2002, pp. 393-422, doi: 10.1016/S1389-1286(01)00302-4.

[3] N. Nasri, A. Gabri, and A. Kachouri, "An Improved DV-Hop Node Algorithm in Wireless Sensor Networks," African Journal of

Information and Communication Technology, vol. 7, issue 2, 2013, pp. 73-82.

[4] N. Nasri, A. Ghabri, S. Khediri, and A. Kachouri, "New Geometric Positioning Approach For Wireless Sensor Networks," The First International Workshop on Performance Evaluation and Modeling in Wireless Networks (PEMWN), Tunis, Tunisia, 22 November 2012, pp. 1-6.

[5] M. Abdallah, J. Bahi, and A. Mostefaoui, "Sensor Networks: localization, coverage and data fusion," Franche-Comté, 14 November 2008.

[6] R. Nagpal, "Organizing a Global Coordinate System from Local Information on an Amorphous Computer," A. I. Memo 1666, MIT A. I. Laboratory, August 1999.

[7] R. Nagpal, H. Shrobe, and J. Bachrach, "Organizing a global coordinate system from local information on an ad hoc sensor network," in proceedings of the 2nd international conference on Information processing in sensor networks, Berlin, 2003, pp. 333-348, doi: 10.1007/3-540-36978-3_22.

[8] L. Gui, "A New Location Algorithm for Wireless Sensor Networks," Ph.D thesis, EDSYS, Toulouse, INSA-Toulouse, 2011.

[9] D. Niculescu and B. Nath, "DV-based positioning in ad hoc Networks," Kluwer Journal of Telecommunication Systems, vol. 22, issue 1-4, January 2003, pp. 267-280, doi: 10.1023/A:1023403323460.

[10] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher "Range-free localization schemes for large scale sensor networks," in proceedings of the 9th Annual international conference on Mobile computing and networking, New York, 2003, pp. 81-95, doi: 10.1145/938985.938995.

[11] T. Gao, D. Greenspan, M. Welsh, and R. Juang, "Vital signs monitoring and patient tracking over a wireless network", in proceedings of 27th Annual International Conference of the Engineering in Medicine and Biology Society, Shanghai, 2005, pp. 102-105, doi: 10.1109/IEMBS.2005.1616352.

[12] H. Baldus, K. Klabunde, and G. Muesch, "Reliable setup of medical body sensor networks," in proceedings of the European conference on Wireless Sensor Networks (EWSN), Wireless sensor networks, vol. 2920, Germany, 2004, pp. 353-363, doi: 10.1007/978-3-540-24606-0_24.

[13] G. Saout, A. Wei, and L. Redon, "Localization in systems of surveillance and support for the elderly," IUT of Blagnac, University of Toulouse, UTM-LATTIS, 1 Place Georges Brassens, F-31703 Blagnac, France, pp. 1-8.

[14] J. Lloret, J. Tomas, M. Garcia, and A. Canovas, "A Hybrid Stochastic Approach for Self-Location of Wireless Sensors in Indoor Environments", Sensors, vol. 9, issue 5, May 2009, pp. 3695-3712, doi: 10.3390/s90503695.

[15] C. Basaran, S. Baydere, and G. Kucuk, "A Hybrid Localization Algorithm for Wireless Sensor Networks," IEICE Transactions on Communications, vol. E91. B, issue 6, 2010, pp. 1852-1861, doi: 10.1093/ietcom/e91-b.6.1852.

[16] D. Niculescu and B. Nath. "Ad Hoc Positioning System (APS)," in proceedings of the IEEE Global Telecommunications Conference, vol. 5, San Antonio, November 2001, pp. 2926-2931, doi: 10.1109/GLOCOM.2001.965964.

[17] X. Chen and B. Zhang, "Improved DV-Hop Node Localization Algorithm inWireless Sensor Networks," International Journal of Distributed Sensor Networks, vol. 2012, July 2012, pp. 1-7, doi: 10.1155/2012/213980.

[18] L. Doherty, K. Pister, and L. E. Ghaoui, "Convex position estimation in wireless sensor networks," in proceedings of IEEE Computer and Communications Societies (INFOCOM), vol. 3, Anchorage, AK, April 2001, pp. 1655-1663, doi: 10.1109/INFCOM.2001.916662.

[19] R. K. Patro, "Localization in wireless sensor network with mobile beacons," in Proceedings of 23rd IEEE Convention of Electrical and Electronics Engineers in Israel, September 2004, pp. 22-24, doi: 10.1109/EEEI.2004.1361078.

[20] M. Essoloh, C. Richard, and H. Snoussi, "Anchor-based distributed localization in wireless sensor networks," 14th Workshop on Statistical Signal Processing, vol. 1, Madison, WI, USA, August 2007, pp. 393-397, doi: 10.1109/10.1109/SSP.2007.4301287.

[21] J. P. Sheu, P. C. Chen, and C. S. Hsu, "A distributed localization scheme for wireless sensor networks with improved grid-scan and vector-based refinement," IEEE Transactions on Mobile Computing, vol. 7, issue 9, June 2008, pp. 1110-1123, doi: 10.1109/TMC.2008.35.

[22] H. Chen, K. Sezaki, P. Deng, and C. Hing, "An improved DV-hop localization algorithm for wireless sensor networks," 3rd IEEE Conference on Industrial Electronics and Applications, Singapore, June 2008, pp. 1557-1561, doi: 10.1109/ICIEA.2008.4582780.

[23] D. Ying, W. Jianping, and Z. Chongwei, "Improvement of DV-Hop Localization Algorithms for Wireless Sensor Networks," 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), Chengdu, September 2010, pp. 1-4, doi: 10.1109/WICOM.2010.5601372.

[24] L. Hui, D. Houshang, B. Pat, and L. Jing, " Survey of Wireless Indoor Positioning Techniques and Systems," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 37, issue 6, November 2007, pp. 1067-1080, doi: 10.1109/TSMCC.2007.905750.

[25] Z. Li, R. Li., Y. Wei, and T. Pei, "Survey of localization techniques in wireless sensor networks," Information Technology Journal, vol. 9, issue 8, September 2010, pp. 1754-1757, doi: 10.3923/itj.2010.1754.1757.

[26] L. Wen-yuan, W. En-shuang, C. Zi-jun, and W. Lin, "An Improved DV-Hop Localization Algorithm based on The Selection of Beacon Nodes," Journal of Convergence Information Technology, vol. 5, issue 9, November 2010, pp. 157-164, doi. 10.4156/jcit.vol5.issue9.16.

A Framework for Supporting Natural Interaction with Printed Matter in Ambient Intelligence Environments

George Margetis, Margherita Antona
Foundation for Research and Technology – Hellas
Heraklion, Greece
e-mail: gmarget@ics.forth.gr, antona@ics.forth.gr

Constantine Stephanidis
Foundation for Research and Technology – Hellas
Heraklion, Greece
Department of Computer Science, University of Crete
Heraklion, Greece
e-mail: cs@ics.forth.gr

***Abstract*—Paper is a widely used material which - through appropriate technological augmentation - has the potential to become a widely accepted means of interaction. Ambient Intelligence bears the promise of smart, adaptive and user-friendly environments, anticipating user needs in an unobtrusive manner. So far, there is no systematic approach to paper augmentation in Ambient Intelligence. Addressing this need, this paper introduces an extensible context-aware interaction framework to enable the integration of printed matter into Ambient Intelligence environments.**

***Keywords*—ambient intelligence; printed matter; interactive paper; natural interaction.**

I. INTRODUCTION

Through the centuries, paper prevailed as the major means for information sharing among people. With the invention of the printing press by Gutenberg, a vast burst of information dissemination occurred all over the world, establishing printed matter as an essential part of people's everyday life. Since the early 90's, the idea of digitally augmenting physical paper was intriguing enough to trigger the first research efforts in this direction. Since then, numerous approaches have been proposed, based on paper's affordances, for providing user interaction. Paper-based interaction has the potential to be widely accepted and applied in everyday life, due to a fundamental prop of paper: it is inexpensive and can be found anywhere.

The recent emergence of Ambient Intelligence (AmI) realizes the vision of a technological environment where the emphasis is on greater user-friendliness, provision of more efficient services, user-empowerment, and support for human interactions. In AmI environments, people are surrounded by intelligent intuitive interfaces that are embedded in all kinds of objects, while the environment is capable of recognizing and responding to the presence of different individuals in a seamless, unobtrusive and often invisible way [1]. AmI has profound consequences on the type, content and functionality of the emerging digital products and services, as well as on the way people interact with them, bringing about multiple new requirements.

Aarts and Marzano in [2] discuss the fundamental features that characterize Ambient Intelligence environments, which can be summarized in five concepts: technology embedment, context awareness, personalization, adaptiveness and anticipation. Although several approaches have contributed frameworks that embed technology in everyday environments, anticipate and address everyday life needs in adaptive and personalized ways, and provide natural interaction with the use of physical or smart objects, there are so far no systematic approaches engaging printed matter towards realising such concepts.

This paper discusses a systematic approach to fill this gap by developing an extensible context-aware interaction framework, which will enable the integration of printed matter into AmI environments, thus providing:

- multimodal natural interaction with printed matter
- printed matter augmentation
- a reference model for printed matter context-aware and anticipation mechanisms, based on a proposed ontology.

For the assessment and evaluation of the proposed framework with end users, four Ambient Intelligence applications are presented, constituting indicative real life examples.

The rest of the paper is organized as follows:

Section II discusses related work, highlighting efforts towards printed matter and paper digital augmentation. Section III presents the overall architecture of the proposed framework. Section IV describes a generic approach for printed matter modelling and profiling. Section V discusses the proposed ontology scheme for enabling context awareness in the use of printed matter in AmI Environments. Section VI focuses on two fundamental modules of the proposed framework that are responsible for the augmentation of printed matter and interaction rendering. Section VII presents four example applications that have been designed and developed using the proposed framework. Finally, Section VIII summarises the paper and highlights next steps and future work.

II. RELATED WORK

The idea of digitally augmenting physical paper was firstly introduced in two pioneer systems, namely DigitalDesk [3] and its successor EnhancedDesk [4], which performed physical paper augmentation offering interaction via touch.

Since then, numerous approaches toward physical paper augmentation emerged, setting the frontier for the interactive paper era. Most of these approaches focused on paper's affordances, such as its light weight and the capability of annotating content , but were also based on the fact that

 ISBN: 978-1-61208-421-3

paper constitutes a fundamental means of information dissemination.

In 1999, Mackay and Favard in [5] introduced the term “interactive paper”, signifying the potential role of digitally augmented paper for the forthcoming technologies. One such example is the Anoto system [6], which combines a unique pattern printed on each page with a digital pen to capture strokes made on paper. PapierCraft [7], on the other hand, is a system which uses pen gestures on paper to support active reading and allows users to carry out a number of actions, such as copying and pasting information from one document to another.

Other approaches to digitally augmented paper use touch and gestures as basic interaction technique. For example, the Pacer system [8] provides gestures support and touch based interaction with printed paper through the touch screen of a cameraphone. Pointing and writing in augmented reality environments has also been studied, but the majority of research work is based on proprietary technological artefacts e.g., light pens, pen with pads, and haptic devices [9] [10].

In terms of visualizing physical paper augmentation, a diversity of different approaches have emerged. For example, MagicBook [11] provides augmentation of physical books with 3D graphics and moving avatars through VR glasses, giving to the reader the sense of living pages. Pacer [8] supports printed paper augmentation via smartphones’ camera, acquiring images of the physical paper in real time and displaying them augmented with digital content on a smartphone’s display. Korozi et al. in [12] present two educational mini-games that offer physical interaction on a tabletop setup through printed cards, where a simple webcam monitors the table's surface and identifies the thrown cards, while the digital content is displayed on a nearby screen. In [13], an interactive desk that augments physical papers placed upon its surface with multimedia content and interactive applications is discussed. This system augments physical paper by projecting the digital content either on the paper or laterally to it.

Although a large number of approaches consider physical paper or printed matter as a fundamental means of interaction with technological artefacts in everyday life, there is still a lack of holistic approaches placing digital paper in the context of Ambient Intelligence in terms of technology embedment, context awareness, personalization, adaptiveness and anticipation. In this paper, an integrated solution for the use of printed matter and physical paper in Ambient Intelligence environments is proposed.

III. THE FRAMEWORK

According to Cook et al. [14], any smart environment can be adequately decomposed in four fundamental layers: physical, communication, information and decision. Each layer performs a different role in the environment, facilitating diverse operations and addressing specific requirements.

The discussed framework has been designed in order to facilitate the development of smart systems that use printed matter or physical paper as a main means of interaction in Ambient Intelligence environments.

Figure 1 illustrates the overall architecture of the proposed framework. Beginning bottom-up, the physical layer of the system comprises the hardware accompanied by the necessary software for printed matter recognition and tracking, as well as the supported user interaction techniques. Since the physical layer can consist of heterogeneous and alternative printed matter recognition systems based on different approaches (e.g., computer vision, electronic markers, etc.), the *Printed matter modeller* provides the digital “alter ego” of the physical subject. Furthermore, the *Annotation tool* provides intuitive UIs for printed matter modelling, through which the developers can make available to the system digital information corresponding to the

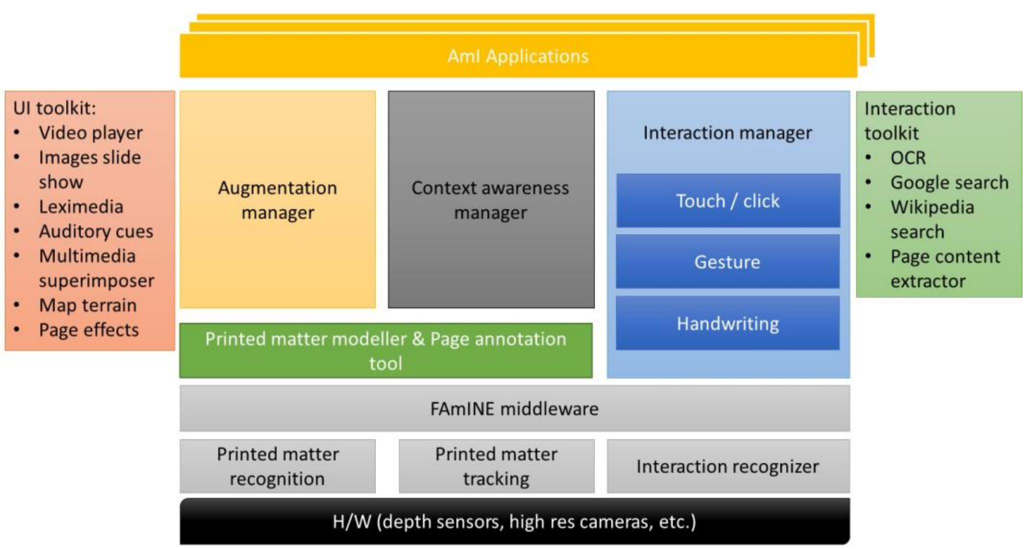


Figure 1. Framework’s architecture

printed matter (e.g., multimedia content, references to internet sources, etc.)

The communication layer is realized by *FAmINE*, a middleware software developed in the context of the FORTH-ICS's AmI Programme and Smart Environments, providing the necessary seamless interoperability of the devices and services that comprise the AmI ecosystem.

The decision layer is implemented by the *Context Awareness Manager*, a fundamental component of the proposed framework, which is responsible for the selection of the appropriate UI components and corresponding content, according to the user's interaction in a specific context of use. The type of interaction is provided by the *Interaction manager*, which undertakes the task of interpreting users' interactions with printed matter. In order for the *Interaction manager* to render the supported types of interaction and for the *Context Manager* to extrapolate the necessary information from the corresponding printed matter, an *Interaction Toolkit* has been implemented including a number of external processes, such as Optical Character Recognition (OCR), information harvesting from various internet sources (e.g., Google search, Wikipedia), and a page content extractor (e.g., extracts text or images from the open pages, etc.)

The information layer consists of the *Augmentation manager*, which is responsible for the rendering of the available UIs provided by the framework for printed matter digital augmentation, adapted to the users' preferences and needs. Moreover, a number of fundamental UI components for printed matter augmentation has been developed and included in the *UI Toolkit*.

Each of the abovementioned components is *printed matter centric*, meaning that they address interaction and augmentation requirements for using printed matter. Furthermore, these components implement the necessary functionality for realizing the fundamental properties of AmI environments. For example, the *Context awareness manager* provides a user / context modelling scheme and ontology-based reasoning enabling personalization, context awareness and anticipation of users' needs. On the other hand, the *Augmentation manager* facilitates adaptivity mechanisms, offering alternative UIs according to the devices where an application is deployed, the profile and preferences of potential users, as well as the type of augmentation supported by an application.

IV. PRINTED MATTER MODEL

In order for the framework to keep structured information about the digital instance of printed matter, the *Printed matter modeller* has been implemented.

This component provides a classification of printed matter in an extended version of the XML description discussed in [24], including the digital representation (e.g. high resolution image) of the printed matter and interactive areas (hotspots) accompanied with their properties, stored in a recognition database. Every single matter in the recognition database is referenced by a unique id and is accompanied by its digital representation path. This digital representation is necessary for printed matter recognition by the framework, but it can also be displayed on any interactive screen near the physical paper or directly on it, using a video projector, enabling therefore the user to interact with hotspots that may be provided.

Every interactive hotspot is declared by a set of coordination points (normalized in order to be independent of the printed matter size), representing the hotspot's bounding path and a set of metadata information regarding the actual content of the hotspot such as the type (e.g.,

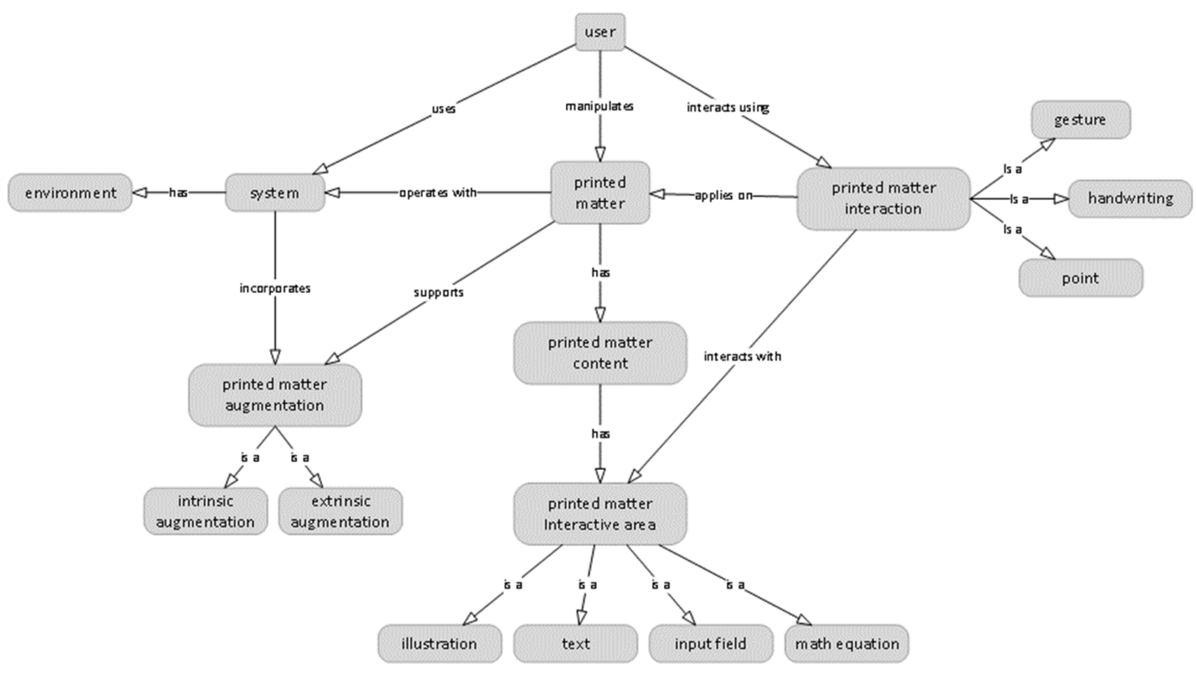


Figure 2. The framework's ontology

image, textual information, input placeholder), a description and a number of keywords.

The aforementioned information can be edited by the developer or the users themselves through the *Page annotation tool.*

The printed matter model constitutes the basic input for the *Context awareness manager* to provide appropriate content to the *Augmentation manager* and to consume interaction events fed by the *Interaction manager.*

V. CONTEXT AWARENESS

As already discussed in the previous section, the proposed framework provides the necessary tools and a software infrastructure for developers to easily integrate paper-based systems in AmI environments. A basic concern in this respect is the support of user modelling and context-awareness. To this end, a new ontology scheme has been designed, illustrated in Figure 2.

For the genericity of the framework's approach, physical paper is considered a subclass of printed matter. It should be noted that white paper doesn't provide any content and therefore it can be incorporated as soon as anything is written or printed on it. Physical paper can be considered as printed matter as soon as something is written or printed on it.

The basic entities of the aforementioned ontology are:

- **user:** The user entity of the proposed ontology captures users' needs in an AmI environment. It can be extended by user attributes or entities, such as users' profile, or their role and potential tasks that they may perform in the context of an environment.
- **printed matter:** It constitutes the generic class that describes the common properties of physical paper or any similar matter that contains handwritten or printed information.
- **printed matter interaction:** It generalizes the potential ways of interaction of the users with printed matter. Children of this class are: (a) gestural interaction (e.g., cycling a paper region, making a pinch gesture), (b) handwriting using a stylus object and (c) pointing / clicking either using fingers or stylus objects.
- **system:** The generic class of AmI stystems, ranging from single smart artefacts to sophisticated platforms (i.e., sophisticated systems comprising multiple sensors, running various services and running on heterogeneous hardware).
- **environment:** The generic class that defines environmental properties. Directly correlated classes are the location of the systems, the time that users' actions are performed and the environmental conditions in terms of temperature, pressure, humidity, lighting and noise.
- **printed matter augmentation:** It is divided in two subclasses (a) intrinsic augmentation that includes all augmentation techniques that apply on the printed matter itself and (b) extrinsic augmentation that regards augmentation techniques, which apply laterally or at a short distance from the printed matter.
- **printed matter content:** The generic class referring to handwritten or printed information on the printed matter. This class can be the generalization of the *document* entity and its subclasses are as defined in [15].
- **printed matter interactive area:** Refers to the types of interactive areas that can be found on printed matter: (a) *illustration* including any type of illustrated picture of figure (e.g., images, graphs), (b) *text* regarding any handwritten or printed textual information, and (c) *input field* referring to any type of printed placeholders that need users' input (e.g., text fields, checkboxes, etc.) and (d) *math equations.*

The proposed ontology can be easily extended using existing ones, for example [16], providing thus an open ontology scheme that can be used for context awareness in AmI environments.

The implementation of the proposed ontology makes use of the Web Ontology Language (OWL) [17]. In conjunction with a reasoning engine (e.g. Apache Jena [18], Jess [19], Microsoft Workflow Foundation Rules Engine [20]), the ontology constitutes the basic component of the *Context manager* module that is responsible for providing context awareness to the applications using the framework.

VI. CONTENT VISUALIZATION AND USER INTERACTION

As already mentioned in Section III, two modules have been implemented for content provision and interaction management, the *Augmentation manager* and the *Interaction manager.*

The *Augmentation manager* handles the visualization of the content in terms of appropriate UIs selection. The key factors that are mainly considered for the visualization of the available content are: (i) whether the provided content will be displayed on the surface of the printed matter or near it, (ii) the visualization output properties (e.g., a projection juxtapose to the printed matter, a nearby display, absence of display), (iii) users' preferences and (iv) the environment (e.g., whether the visualization applies in a noisy or silent environment).

The selection of the appropriate UI is made from a set of basic UIs that have been especially designed in order to address the needs of visualization of appropriate content in the context of printed matter manipulation in AmI environments. The UI components that have been implemented are able to visualize and provide interaction with heterogeneous content, including images, videos, textual information, geospatial information, auditory cues, animations and effects on the physical printed matter or digital representation of it, as well as any combination of these components.

On the other hand, the *Interaction manager* processes the input of several interaction techniques that the framework provides, such as touch, gestures and handwriting on printed matter or on a provided UI. According to the type of interaction, different types of information are provided. For example, if a user points at a hotspot area of a printed matter, then the *Interaction manager* will provide only the corresponding metadata information to the *Context awareness manager* for further processing. On the other hand, if a user makes a rectangular gesture denoting that the designated area should be isolated for further processing (e.g., annotation by the user), the *Interaction manager* extracts this area in the form of a digital image and also acquires possible referenced information about this area such as the text that it may contain (extracted either from the description of the printed matter model or directly using Online Character Recognition). This information is also provided to the *Context manager* for further elaboration.

Along with the *Interaction manager*, a toolkit is provided containing tools for data extraction from the printed matter and semantic information search from online web services.

VII. APPLICATIONS

For the assessment of the proposed framework four example applications have been developed. These applications were evaluated in terms of usability and user experience according to standard evaluation procedures [17] [22] [23].

The first application (SESIL) [24] is an educational system that incorporates the proposed framework in order to provide stylus-based interaction in different spatial arrangements, such as large interactive surfaces featuring a display with multi-touch capabilities (i.e., for use in a library or at an exposition) using cameras for stylus recognition and tracking.

The system aims at enhancing reading and writing activities on physical books through unobtrusive monitoring of users' gestures and handwriting, as well as the display of information related to the current users' focus of attention. Additionally, it exploits the *Context manager* to decide at run-time the type of additional information and support to be provided in a context-dependent fashion.

The system consists of a desk with a set of three high resolution cameras placed above it to achieve the recognition and tracking of the school books placed on the desk's surface. A nearby large display runs an educational application that provides content-sensitive information to the users, based on their stylus-based interaction with a school book, following the extrinsic type of printed matter augmentation.

The second application (Book of Ellie) [25] is the augmented version of a classic schoolbook for teaching the Greek alphabet to primary school children. The book introduces alphabet letters and their combinations by increasing the difficulty level. For each letter or letter combination, relevant images and text involving the specific letter(s) are provided. The short stories for each letter are structured around dialogues and activities of a typical Greek family, with the protagonist being Ellie, one of the four children. In the augmented version of the book, Ellie has become an animated character, constantly available to assist the young learner by reading phrases from the book, asking questions or providing advice.

In terms of setup, the system consists of a television screen (32'') for visual and audio output, an "Asus Xtion Pro" RGBD camera, and a PC running the software. The RGBD camera is used to recognize and localize book pages and cards, as well as to detect and localize fingertip contacts on the book and table. The physical book and paper cards (e.g., depicting letters, simple objects, or animals) are interactive components of the system.

The third application is an Augmented Reality (AR) study desk [13] [24], which aims at augmenting physical books with digital information. The system consists of a standard definition projector and an ASUS Xtion Pro, both overlooking the surface of a desk. The images acquired by the color camera of the Xtion are used for printed matter recognition and its localization on the desk surface, while the images acquired by the Xtion's depth camera are used for detecting users' finger touch on the printed matter or the desk.

The AR study desk provides context-aware multimedia and interactive applications related to the content of the open book page. Such content is dynamically displayed to enrich the contents of the currently open book page, and is aligned, in real-time, with its 2D orientation upon the desk.

Technically, augmentation is supported by the projector-camera calibration. Given the coordinates of the book or the stylus in the desk coordinate frame, this calibration is used to predict the coordinates of the projector pixel that will illuminate the corresponding region or point of interest.

The last application (Study Buddy) [26] provides an unobtrusive intelligent environment that implements a context aware system targeted to augment the learning process. The system is composed of a smart reading lamp and educational software, called LexiMedia, aiming to provide dictionary information, as well as multimedia information for specific words, thus assisting in language learning.

The smart reading lamp incorporates a small camera and an embedded computer with WiFi connection. The camera of the reading lamp targets to the student's reading area (i.e., the area of the desk where the book is placed). Interaction with the system is initiated when a user indicates a word in the book, by using a black pointer (e.g., pen) and carrying out one of the following gestures: pointing at the word, underlining the word or circling the word.

Whenever the smart reading lamp observes that the reader needs help about a word or a phrase, it scans the area trying to recognize the indicated words, using OCR software. Then, it collects useful information about the recognized words, such as related images and words' definition. Finally, it transmits the aforementioned information to device (e.g., tablet, smart phone, etc.) which runs LexiMedia, placed near the reader .

TABLE I summarizes the interaction modalities and augmentation features that the aforementioned applications provide using the components of the discussed framework.

TABLE I. SUMMARY OF INTERACTION MODALITIES AND DIGITAL AUGMENTATION PROVIDED BY THE APPLICATIONS

Application	Interaction modalities	Augmentation
SESIL	• Handwriting • Stylus gestures • Page flipping • Recognition and localization of books on desk	Provides context-sensitive assistive content to the students on a nearby display.

Application	Interaction modalities	Augmentation
Book of Ellie	• Hand gestures • Page flipping • Printed cards position and orientation recognition • Recognition and localization of books on desk	Provides auditory cues to the students on a nearby device. Provides card-based educational games of varying difficulty according to the students' profile.
Study desk	• Hand gestures • Page flipping • Recognition and localization of books on desk	Provides context-sensitive content on top of or latteraly to the open page.
Study buddy	• Handwriting • Stylus gestures • Page flipping • Recognition and localization of books on desk	Provides thesaurus information and explanatory multimedia content for any english word that can be found on the open page.

VIII. CONCLUSIONS

Taking into account the potential of physical paper to become a widely accepted interactive component of an Ambient Intelligence environment, an extensible context-aware interaction framework has been proposed, which enables the integration of printed matter into AmI environments. The proposed framework allows application developers to enrich printed matter with digital information according to their application's envisioned context of use, supports natural multimodal interaction both with physical paper and the digital content resulting from the paper augmentation, and features multimedia output.

Future work will address the issue of multimodal interaction with printed matter in combination with other physical objects coexisting in the environment. Furthermore, the generalization of the proposed framework will be investigated towards incorporating other physical objects. Another aspect to be addressed is the investigation of automated recognition and modelling of printed matter based on its content web semantic analysis. Finally, future efforts will include the evaluation of the framework by software developers of AmI applications.

AKNOWLEDGMENT

This work was supported by the Foundation for Research and Technology – Hellas, Institute of Computer Science (FORTH – ICS) internal RTD Programme 'Ambient Intelligence and Smart Environments'.

REFERENCES

[1] IST Advisory Group, Scenarios for Ambient Intelligence in 2010. [Online]. Available from: ftp://ftp.cordis.lu/pub/ist/docs/istagscenarios2010.pdf [retrieved: June, 2015].

[2] E. H. Aarts and S. Marzano, (Eds.). "The new everyday: Views on ambient intelligence", 010 Publishers, 2003.

[3] P. Wellner, "Interacting with paper on the DigitalDesk", Commun. ACM, vol. 36(7), pp. 87–96, 1993.

[4] M. Kobayashi and H. Koike "EnhancedDesk: Integrating paper documents and digital documents", Asia Pacific Computer Human Interaction (APCHI'98), July 1998, pp. 57–62.

[5] W. E. Mackay and A. L. Fayard, "Designing interactive paper: lessons from three augmented reality projects", Proc. International workshop on Augmented reality: placing artificial objects in real scenes, AK Peters, Ltd., 1999, pp. 81-90.

[6] Anoto, Development Guide for Service Enabled by Anoto Functionality, 2002.

[7] C. Liao, F. Guimbreti, K. Hinckley, and J. Hollan, "Papiercraft: A gesture-based command system for interactive paper", ACM Transactions in Computer-Human Interactions, vol. 14 (4), pp. 18 – 55, 2008.

[8] C. Liao, Q. Liu, B. Liew, and L. Wilcox, "Pacer: fine-grained interactive paper via camera-touch hybrid gestures on a cell phone", Proc. SIGCHI Conference on Human Factors in Computing Systems, ACM, 2010, pp. 2441-2450.

[9] A. S. Forsberg, Jr. J. J. LaViola, and R. C. Zeleznik, "ErgoDesk: A Framework for Two- and Three-Dimensional Interaction at the ActiveDesk", Second International Immersive Projection Technology Workshop, May 998, pp. 11-12.

[10] D. Schmalstieg, et al., "The studierstube augmented reality project", Presence: Teleoper. Virtual Environ., vol. 11(1), pp. 33-54, 2002.

[11] M. Billinghurst, H. Kato, and I. Poupyrev, "The MagicBook: Moving Seamlessly between Reality and Virtuality", IEEE Computer Graphics and Applications, vol. 21(3), pp. 6-8, 2001.

[12] M. Korozi, et al., "Ambient educational mini-games", Proc. International Working Conference on Advanced Visual Interfaces, ACM, 2012, pp. 802-803.

[13] G. Margetis, A. Ntelidakis, X. Zabulis, S. Ntoa, P. Koutlemanis, and C. Stephanidis, "Augmenting physical books towards education enhancement", 1st IEEE Workshop on User-Centered Computer Vision (UCCV 2013), IEEE, Jan. 2013, pp. 43-49.

[14] D. J. Cook and S. K. Das, "How smart are our environments? An updated look at the state of the art", Pervasive and Mobile Computing, vol. 3(2), pp. 53–73, 2007.

[15] Bibliographic Ontology Specification. [Online]. Available from: http://bibliontology.com/ [retrieved: June, 2015]

[16] D. Preuveneers, et al., "Towards an extensible context ontology for ambient intelligence", Proc. Second European Symposium Ambient intelligence (EUSAI 2004), Springer Berlin Heidelberg, Nov. 2004, pp. 148-159.

[17] OWL Web Ontology Language Overview. [Online]. Available from: http://www.w3.org/TR/2004/REC-owl-features-20040210/#s1.1 [retrieved: June, 2015]

[18] Jena documentation overview. [Online]. Available from: https://jena.apache.org/documentation/index.html [retrieved: June, 2015]

[19] Jess, the Rule Engine for the Java Platform. [Online]. Available from: http://www.jessrules.com [retrieved: June, 2015]

[20] Introduction to the Windows Workflow Foundation Rules Engine. [Online]. Available from: https://msdn.microsoft.com/en-us/library/aa480193.aspx [retrieved: June, 2015]

[21] J. Nielsen, "Usability engineering", Morgan Kaufmann, pp. 165-206, 1994.

[22] J. Nielsen, "Usability engineering", Morgan Kaufmann, pp. 207-226, 1994.

[23] C. Lewis, “Using the ‘thinking-aloud’ method in cognitive interface design”, IBM Research Report RC 9265, Yorktown Heights, NY, 1982.

[24] G. Margetis, X. Zabulis, P. Koutlemanis, M. Antona, and C. Stephanidis, “Augmented interaction with physical books in an Ambient Intelligence learning environment”, Multimedia Tools and Applications, vol 67(2), pp. 473-495, Jan. 2012.

[25] E. Papadaki, et al., “The Book of Ellie: An Interactive Book for Teaching the Alphabet to Children”, International Conference on Multimedia and Expo Workshops (ICMEW), IEEE, July 2013, pp. 1-6.

[26] G. Margetis, S. Ntoa, M. Bouhli, and C. Stephanidis, “Study-Buddy: Improving the Learning Process through Technology-Augmented Studying Environments”, HCI International 2011–Posters’ Extended Abstracts, Springer Berlin Heidelberg, 2011, pp. 504-508.

Multi-Target Data Association in Binary Sensor Networks

Sebastian Matthias Müller
OFFIS – Institute for Information Technology
Oldenburg, Germany
Email: sebastian.mueller@offis.de

Andreas Hein
School of Medicine and Health Sciences
University of Oldenburg
Oldenburg, Germany
Email: andreas.hein@uni-oldenburg.de

***Abstract*—Numerous applications of ambulant medical care, house automation and security use binary sensors such as passive infrared motion sensors or light barriers to monitor activity in the house. Multi-target tracking algorithms allow for at least a partial separation of activity in data from such sensors from multiple persons. While many tracking algorithms demonstrate good performance across various sensing modalities and sensor setups, little research has been done to determine the impact of placement and varying density of sensors for tracking performance. This paper presents the results of an evaluation of a Bayesian multi-hypothesis multi-target tracking algorithm on data of two residents monitored by a network of binary sensors. We evaluate the algorithm on data from sensors of varying quantity and placement. We show that our approach outperforms other approaches in low-resolution setups. While tracking performance naturally decreases with the number of sensors, it also strongly varies by sensor positioning.**

***Keywords*–Multi-target tracking; Assisted living; Wireless sensor networks.**

I. Introduction

The emergence of research on technical support systems for ambulant care and support for patients and elderly stem from numerous recent societal developments as well as changes in demographic structure.

First, the coincidence of prolonged life expectancy [1] and the atomization of households [2] puts an increasing care demand into the hands of third parties. According to the German Federal Statistical Office, the number of single-households will increase sixfold in relation to the population numbers. At the same time, the ratio between care personnel supply and demand will cut in half [3].

Second, the increasing life expectancy, in combination with improved medical care and "modern lifestyles and behavior" [4] causes an increase in the proportion of population living the chronic diseases, thus further driving demand for ambulant care.

Third, there is a general trend towards outpatient care by hospitals. According to the Avalere Health analysis of American Hospital Association Annual Survey [5], the percentage of revenue for community hospitals in the United States has increased from 25% to 44% between 1992 and 2012.

These developments drive the research on technical support systems in home and care environments. Applications for such include automated assessments [6], activity monitoring [7] or fall detection [8]. To preserve a maximum of privacy and comfort while at the same time collecting data necessary for the application, many approaches include the use of ambient sensors such as motion sensors and light barriers. Since the data collected from these sensors does not carry identifying information, use of any such application in settings where more than one person – the patient – moves or resides becomes difficult.

Complex sensors, such as cameras and microphones are rarely accepted in living spaces. Body-worn sensors are often forgotten or ignored due to discomfort. Binary sensors such as light barriers and motion sensors are easy to retrofit, have relatively little power consumption and can be installed unobtrusively. A no-requirements sensor model also enables us to install more complex sensors (such as laser scanners or depth-finding cameras) as required. The necessary information can be extracted from their data by partitioning the sensors' range and converting activity in each partition to a binary signal.

To separate data from multiple persons moving in a space monitored by binary sensors, we present a multi-target tracking algorithm using Bayesian estimation and multi-hypothesis tracking. This algorithm makes no assumptions on the selection and placement of sensors or sensing technology. Tracking takes place on a graph of the sensors and their spatial relation. It is thus not helpful in determining the precise location of a present person, but at (or below) room-level accuracy. This algorithm performs particularly well on low-resolution data, such as when only few binary sensors are used. We test the algorithm across various sets of sensors, varying by placement and number. A decreasing number of sensors will likely have an impact on the tracking accuracy, but is important in regard to energy consumption, costs and user acceptance. We show that data from two residents in an apartment can be separated with high (>90%) accuracy, and that the selection and placement of sensors can play a significant role in tracking accuracy.

The remainder of this article is structured as follows: Section II summarizes related works on multi-target tracking and activity monitoring in the home using binary sensors. Section III describes the theoretical principles surrounding data association and multi-hypothesis tracking for single- and multi-target tracking. Section IV explains how the approach was evaluated, including data preparation, the evaluation function as well as the sensor placement concept. The results of the evaluation are presented in section V. Section VI concludes the article.

II. Related Work

Prior work has shown that data collected from sensor networks allow for the deduction of information used in activity monitoring, care assessments and behavior modeling. Target tracking, in particular multi-target tracking, is a task often applied to visual data such as video feeds and images. The practical application of multi-target tracking in binary or low-resolution home sensor networks has been to little research.

A. Target Tracking in Home Sensor Networks

Wilson and Atkeson [9] describe an algorithm for tracking of multiple persons and their activity status in a binary sensor network. In this work, the authors use a transition matrix representing transition probabilities between sensors. By keeping track of the targets' identities, personal motion models emerge. The data association is achieved using a particle filter. During a five-day experiment in a house instrumented with 49 sensors (contact switches, motion sensors), data during two-person scenarios was correctly assigned 82.1% of the time.

Krüger et al. [10] use a particle filter and *action plans* to assign sensor events from motion sensors and light switches to tracks and simultaneously identify the target. Action plans describe action sequences in terms of sensor data. These plans can be synthesized or learned from historic data. For the evaluation, an office corridor was equipped with six light switches and six motion sensors. The mean squared error across time and all targets is reported as approximately 0.26 for two-person scenarios. The work shows how – similar to trained motion models – previous knowledge of a person's plans can help tracking individuals in binary sensor networks.

Oh and Sastry [11] perform tracking on data of binary sensor networks and passage connectivity graphs. The graphs are calculated from transition probability matrices. A tracking algorithm, derived from the Viterbi algorithm, pruning strategies and multiple target tracking extensions are presented. No evaluation on real world data is conducted.

Marinakis et al. [12] derive the topology of a sensor network in terms of transition times and probabilities from data of unspecified sensors. The authors use Monte Carlo Expectation Maximization to assign activity to agents (people present) in order to build a graph of the sensor network. 95% of the topology of simulated node graphs is recovered correctly. The results for a trial using a network of cameras and photocell-based sensors are not reported.

B. Activity Monitoring in Home Sensor Networks

Numerous studies show that data collected from sensor networks in living spaces allow for the deduction of information relevant in applications of activity monitoring, care assessments and behavior modeling.

Logan et al. showed that ambient motion-based sensors provide the most useful information for detection and classification of daily in-home activities in a study compared to RFID, on-body and on-object sensors. In their study, infrared motion sensors yielded the best results overall, although classification performance on this data was better on activities that are strongly correlated with locations in the home, such as "watching TV" and "meal preparation" [13].

Data from binary sensors can also be used to calculate average room residence time and frequency: Assessment tests are partly realizable by using recordings from light barriers and reed contacts alone [14]. The authors argue that light barriers alone do not constitute sufficient evidence of a person entering a room, because people may change directions between rooms. It is suggested to combine light barriers with sensors covering larger areas. Room residence times are calculated by manually labeling the sensors constituting a room using a floor plan and knowledge of the sensors placements. In a similar study, the authors model user behavior of a resident from the probability of location at a certain time of day and the frequency of presence in a location in a defined period of time [15]. Models are created for rooms individually (bathroom, bedroom, living room, kitchen). Based on the number of anomalous behavior detected, the authors conclude that the models' performance varies by room: Presence in the bathroom is best modeled duration-based, while the timeslot-based model yielded better results for the other rooms.

Frenken et al. [6] use ambient sensors in an attempt to automate measurement of mobility and gait velocity, as required in the Timed Up and Go assessment [16]. For this, five flats are equipped with home automation sensors and one with an additional laser range scanner. It is shown that the data is suitable to compute gait velocity at home. While data from the laser range scanner is proven to be more precise than home automation sensor data, no statistical post-processing or filtering was performed on the latter.

III. Approach

We define a sensor graph of sensors $s_1, \cdots, s_N$ as a weighted, directed graph $G = (V, L)$, where $V = \{1, \cdots, N\}$ is the set of nodes in the graph representing the sensors, and L is the set of all edges (u, v) for which there is a direct passage from the sensing region of sensor u to the sensing region of v which does not intersect any other sensing regions. Informally, two sensors u, v are connected if it is possible for a person to traverse from the sensing region of u to the sensing region of v without activating any other sensor.

Each resident in the target space is represented by a discrete Bayesian filter on an unweighted, undirected graph consisting of sensors as nodes and edges representing their spatial adjacency. For our evaluation data, this graph was published by Crandall et al. [17] (Figure 1). If the adjacency relations are not known, they can be approximated by a path planning algorithm [18] using a floor plan, if available, or generated from historic data [19].

A. Tracking of individuals

Bayesian filters estimate the state of a dynamic system from noisy data. We choose a probability distribution to represent the location of each individual, because it helps estimating a more precise location later on, especially when sensor regions overlap. More importantly still, it helps the tracker to recover more quickly when a noisy measurement is assigned to the individual's track. Lastly, we aim to replace the manually constructed, unweighted graph with a weighted graph that is automatically constructed from in-situ recorded data and transition probabilities between sensors as weights (cf. [12]).

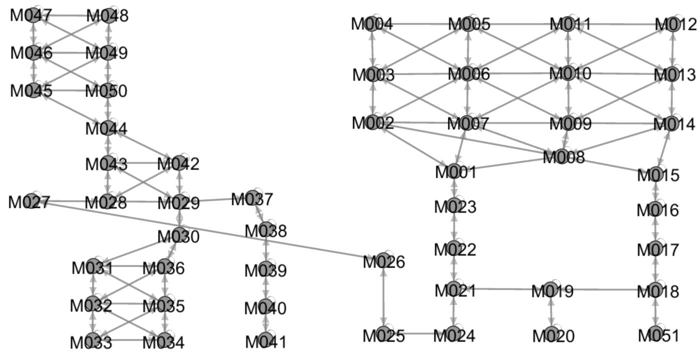


Figure 1. Graph of sensors (with their internal IDs and their spatial relations used in the evaluation (adapted from [17]).

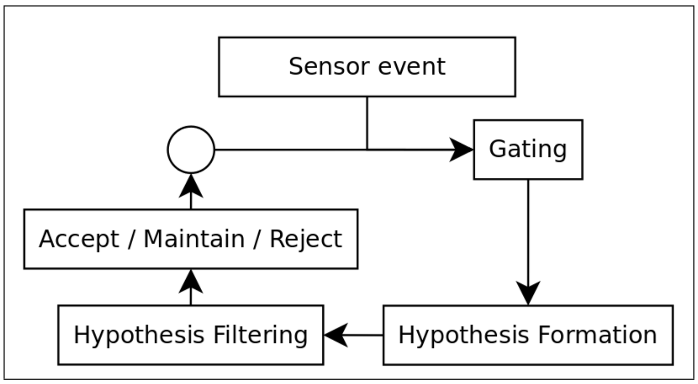


Figure 2. Hypothesis formation overview.

All sensors are subject to noise, but many motion and home automation sensors come with an additional source for noise: measurement delay. Many sensors do not measure or report measurements for a specified amount of time after triggering. This period can last from a few seconds to several minutes due to radio communication regulations. This results in sensors missing the presence or movement, thus breaking the continuity of measurements of a motion track.

At which point and how many filters are created – that is to say, how many individuals are assumed to be present – depends on the performance (*belief*) of the previously existing track: When new measurements cause the current data to be more likely when assigned to more or fewer tracks (= individuals) than before, a new filter is spawned or an existing filter is discarded. The data could be bundled into larger updates within reasonable time frames (cf. [20]), but in our case an update occurs for each new sensor event.

B. Multi-target tracking

When new sensor data arrives, hypotheses are created by considering all possible assignments of the data to existing and new tracks (”hypotheses”) until the filter’s *window size* is reached. This is particularly useful in a low-resolution setting like ours, where individuals may occlude each other in sensor readings for any period of time.

The window size in multi-hypothesis tracking (MHT) describes the maximum number of events (or time steps) that are considered before choosing a likely hypothesis. Windowing is necessary to limit the number of possible hypotheses and to limit the information loss in case no acceptable hypothesis remains and the data is discarded. The influence of the window size on tracking accuracy has been shown previously [21]. For our evaluation, we use a window size of 10 events.

The idea of multi-hypothesis tracking dates back to 1979, when Donald B. Reid published ”An algorithm for tracking multiple targets” [22]. Reid’s algorithm was developed to work on data from a continuous scale sensor (e.g., radar). Therefore, Reid speaks of associating measurements to clusters. In the work presented here, the target space is discrete (nodes on a graph), and targets and their locations are stored as a probability distribution over the space using Bayesian filters.

There are several significant differences between Reid’s original work and the approach described here. In accordance with Reid’s *type 2* sensor, our sensor model expects *positive reports* only, meaning that we consider only sensor data reporting activity. However, tracks are updated per hypothesis, rather than generated and filtered individually (*hypothesis-oriented MHT*). This means that hypotheses are not constructed from *compatible* tracks, but all possible combinations of updates of existing hypotheses. Furthermore, the tracker is updated every time a sensor reports activity. Because of this, and the fact that our state space is discrete, computational complexity is reduced. For a more detailed description of track- and hypothesis-oriented MHT, see Blackman [20].

For each triggered sensor, a new hypothesis based on all previously existing hypotheses is created, in which the triggered sensor is

- considered noise and discarded,
- used to update one of the existing filters, or
- assigned to a new filter.

Due to the exponential growth of the number of possible hypotheses ($> 4.74 \times 10^{13}$ for 20 events), we must employ a number of filters to optimize computation efficiency.

All hypotheses must pass a gating function before they are considered for evaluation (see Figure 2). In our case, this gating function is a simple comparison of the prior probability of each filter to a threshold value. Afterwards, hypotheses are filtered based on confidence, noise ratio and similarity. This procedure is performed until a single hypothesis remains or the window size is reached. In the former case, the hypothesis is accepted, the underlying Bayesian filters updated, and the window size reset. In the latter case, all hypotheses are evaluated. If no single, dominating hypothesis can be found, all hypotheses are discarded and the underlying filters reset.

The size of the window strongly influences the performance of the algorithm. A larger window size will result in a larger number of correct associations, but also in a larger number of discarded sensor events [21].

Figure 2 depicts the general multi-hypothesis tracking logic. For a more in-depth description of multi-hypothesis tracking, see Blackman [20] or Reid [22].

IV. Evaluation

A. Data Preparation

The data used for this evaluation was recorded at the Center for Advanced Studies in Adaptive Systems (CASAS) at the University of Washington [23]. It shows activity of two residents of a smart home environment, residing in a 4-room, 2-story apartment for approximately 8 months. For our evaluation, we use subsets of the data recorded by the 50 motion sensors mounted to the ceiling. The smart home is also equipped with contact sensors on doors and cabinets, temperature, water and electricity sensors. For our purposes, however, motion sensors offer the most precise and least noisy data.

We use data for which at least both residents are present and active. We choose time frames

- that last at least 20 minutes or contain at least 300 sensor events,
- in which both residents change rooms at least once, and
- in which neither resident is inactive for more than 20% of the time.

The result are twenty time frames, with 330 to 910 sensor events with durations between 24 and 530 minutes. After selection, each of the 13321 sensor events was labelled as originating from Resident 1, Resident 2 or a third person using the manually labelled events and the laboratory's floor plan.

B. Data Association

The algorithm can track any number of targets. However, our intended area of application – small households – allows us to use an evaluation function that is tailored towards few targets (1-3). For this evaluation, the algorithm was optimized to track two targets by using an evaluation function that favors one- and two-track hypotheses. Equation (1) describes the evaluation function, where h is the hypothesis in question, $conf(p_n)$ is the belief of the Bayesian filter at the most recent event location n, $\|p\|$ is the number of paths (= targets) in h, and m is the expected number of targets in the sensor space.

$$eval(h) = \frac{\sum_{i=1}^{n} conf(p_n)}{\frac{\|p\|^2+m}{m+1}} \tag{1}$$

C. Sensor Placement

To get a better understanding of how the number of sensors affects tracking accuracy, we also run the algorithm on subsets of the original set of sensors in decreasing size (40, 30 and 20 sensors). Instead of choosing the sensors randomly, we chose characteristics of sensors we deemed possibly influential on tracking performance:

1) Number of neighboring sensors: Based on the assumption that sensors in doorways, which usually have few neighboring sensors, are critical in tracking room transitions, we remove those in larger areas with many neighboring sensors. The number of neighboring sensors can be calculated from the sensor graph.

2) Duration of stay: Given that tracking stationary targets is much simpler than moving targets, we consider subsets of sensors that cover areas in which the average duration of stay is short. The duration of stay can be calculated from the duration between consecutive sensor events in recorded data.

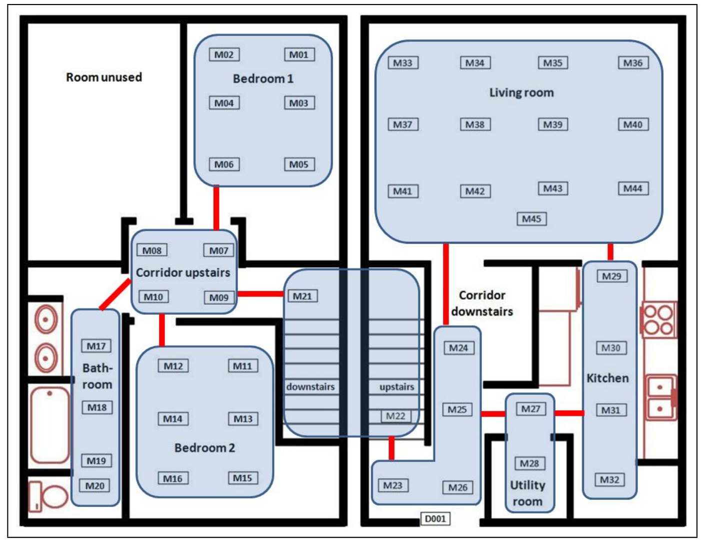


Figure 3. Sensors clustered to represent rooms.

3) Activity: Considering the application of in-home activity monitoring, it is imperative that the placement of sensors for tracking accuracy improvement does not interfere with the necessity of covering those areas in which the majority of activity is taking place. Thus, we select and filter sensors based on the amount of activity covered. The amount of activity covered by a sensor is simply calculated by the number of times it is triggered.

These criteria were used to create subsets of data of varying size, selected by increasing, as well as decreasing order of the respective criterium (cf. Figure 4).

D. Sensor Clustering

The procedure of selecting subsets of sensors for tracking performance evaluation was also conducted for sets of 10 sensors. However, due to the selection criteria, most of the sets had removed whole rooms, and in one case all data from one individual. Thus, in order to evaluate tracking performance on 10 sensors, we cluster the sensors by rooms and spatial adjacency (see Figure 3), and treat the resulting clusters as individual sensors. This also results in a more realistic scenario, in which motion sensors often cover different size areas up to whole rooms.

For this evaluation, we use data from all sensors, but we replace the sensor IDs with IDs for their corresponding cluster. This way, we make use of all sensor events but decrease their spatial resolution.

V. RESULTS

Tracking accuracy using all sensors is 90.3%. This is the percentage of the 13321 sensor events across all time frames that are correctly associated to any of the targets. The accuracy of individual time frames ranges from 62.1% to 99.5%, with a median of 93.1%. The error is composed of false associations (events that are falsely associated to another target, median 5.88%), no associations (events that could not be associated with any target, 0.59%) and noise (events that are falsely discarded as noise, 0.44%).

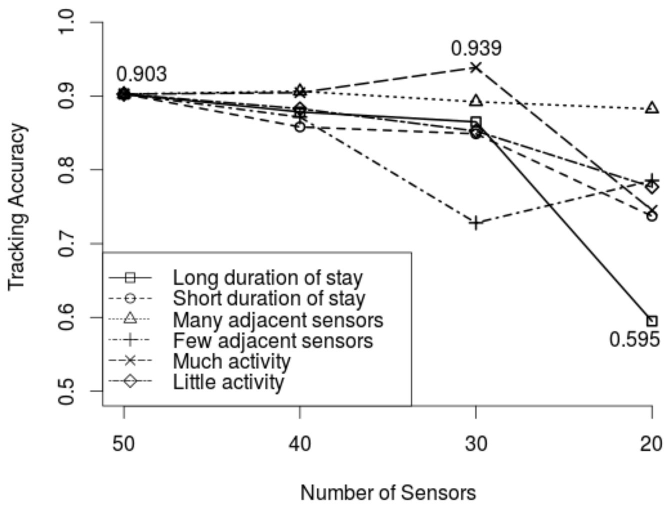


Figure 4. Tracking performance across sensor subsets by size.

Figure 4 shows tracking accuracy across sensor subsets. The sets vary by the number of included sensors (x-axis) and their selection criterion. As can be seen, tracking accuracy generally decreases with reduced sensor count. This is to be expected as the resolution of the tracking space decreases and situations with overlapping motion increases. Down to 30 sensors, tracking accuracy decreases only slightly for all but one sensor set. For the set of sensors with much overall activity, accuracy even increases slightly. The graph also shows that performance variation increases with the number of sensors. While tracking accuracy varies between 85.8% and 90.6% with 40 sensors, with 20 sensors accuracy ranges between 59.5% and 88.2%.

Tracking accuracy on the clustered data set is 77.2%.

VI. CONCLUSION

The article at hand describes an algorithm for tracking of multiple targets in a space monitored by binary sensors. It enables the separation of sensor data generated by multiple persons in smart home environments without the need for identifying sensors. The algorithm makes use of a graph consisting of sensors as nodes and their spatial relations as vertices. Compared to other related works, the algorithm works particularly well in low-resolution settings (i.e., with few binary sensors). It was shown that tracking accuracy can be improved by placing sensors based on activity characteristics. For example, sensors with many neighboring sensors provide a consistently higher accuracy than those with few, and sensors in places where the duration of stay is long on average prove to be less beneficial than those where duration of stay is short.

The data suggests that the decrease of tracking accuracy resulting from smaller sets of sensors (i.e., decreased target space resolution) can be largely absorbed by selective placement of sensors. It was shown that tracking two targets in a network of 20 or more can be achieved for over 90% of the time. The algorithm tracked correctly on ten clusters of motion sensors 77.1% of the time.

It must be noted that differences in tracking performance may not only be due to advantageous sensor placement, but also due to favorable data: While tracking in space with many adjacent sensors works well, it neglects in part space where tracking might be particularly difficult but useful, such as in narrow hallways. The share of total events covered by the different subsets of sensors range from 11 to 98%.

The experiment presented here gives insight into the importance of sensor placement for multi-target tracking using binary sensors. The next step will be to find the ideal sensor setup for the data used in this evaluation, which may be a mixture of the sensor subsets and criteria examined here. Furthermore, the algorithm's performance with more than two targets must be evaluated.

It is further planned to include identifying information in the algorithm so as to not only associate the data to tracks, but to identify the target. This way, the sensor graph can be replaced by an individual motion model, further improving tracking accuracy.

ACKNOWLEDGMENT

The authors would like to thank the team of the Center for Advanced Studies in Adaptive Systems (CASAS) at the Washington State University for making their data publicly available.

This work was funded by the German Ministry for Education and Research (BMBF) within the research project *Cicely* (grant 16SV5896).

REFERENCES

[1] J. Vaupel and H. Lundstrom, "The future of mortality at older ages in developed countries." 1994.

[2] Eisenmenger, M. and Pötzsch, O. and Sommer, B., "Germany's population by 2050, Results of the 11th coordinated population projection," Federal Statistical Office, Tech. Rep., 2006.

[3] H. Rothgang, R. Müller, and R. Unger, Themenreport "Pflege 2030": was ist zu erwarten - was ist zu tun? Bertelsmann Stiftung, 2012.

[4] W. H. Organization et al., "The european health report 2009: Health and health systems," 2009.

[5] A. H. Association et al., The AHA annual survey database. American Hospital Association, 2005.

[6] T. Frenken, B. Vester, M. Brell, and A. Hein, "atug: Fully-automated timed up and go assessment using ambient sensor technologies," in Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on. IEEE, 2011, pp. 55–62.

[7] A. Fleury, M. Vacher, and N. Noury, "Svm-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results," Information Technology in Biomedicine, IEEE Transactions on, vol. 14, no. 2, 2010, pp. 274–283.

[8] A. Sixsmith, N. Johnson, and R. Whatmore, "Pyroelectric ir sensor arrays for fall detection in the older population," in Journal de Physique IV (Proceedings), vol. 128. EDP sciences, 2005, pp. 153–160.

[9] D. Wilson and C. Atkeson, "Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors," in Pervasive computing. Springer, 2005, pp. 62–79.

[10] F. Krüger, M. Kasparick, T. Mundt, and T. Kirste, "Where are my colleagues and why? tracking multiple persons in indoor environments," in Intelligent Environments, International Conference on. IEEE, 2014, pp. 190–197.

[11] S. Oh and S. Sastry, "Tracking on a graph," in Proceedings of the 4th international symposium on Information processing in sensor networks. IEEE Press, 2005, p. 26.

[12] D. Marinakis, G. Dudek, and D. Fleet, "Learning sensor network topology through monte carlo expectation maximization," in Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on. IEEE, 2005, pp. 4581–4587.

[13] B. Logan, J. Healey, M. Philipose, E. Tapia, and S. Intille, "A long-term evaluation of sensing modalities for activity recognition," in UbiComp 2007: Ubiquitous Computing. Springer, 2007, pp. 483–500.

[14] E.-E. Steen, T. Frenken, M. Frenken, and A. Hein, "Functional assessment in elderlies homes: Early results from a field trial," in Ambient Assisted Living. Springer, 2014, pp. 3–17.

[15] E.-E. Steen, T. Frenken, M. Eichelberg, M. Frenken, and A. Hein, "Modeling individual healthy behavior using home automation sensor data: Results from a field trial," Journal of Ambient Intelligence and Smart Environments, vol. 5, no. 5, 2013, pp. 503–523.

[16] D. Podsiadlo and S. Richardson, "The timed" up & go": a test of basic functional mobility for frail elderly persons." Journal of the American geriatrics Society, vol. 39, no. 2, 1991, pp. 142–148.

[17] A. Crandall and D. Cook, Human Behavior Recognition Technologies, H. W. Guesgen and S. Marsland, Eds. IGI Global, Mar. 2013.

[18] T. Frenken, E.-E. Steen, M. Brell, W. Nebel, and A. Hein, "Motion pattern generation and recognition for mobility assessments in domestic environments." in Proceedings of the 1st Int. Living Usability Lab Workshop on AAL Latest Solutions, 2011, pp. 3–12.

[19] S. Müller, A. Helmer, E.-E. Steen, M. Frenken, and A. Hein, "Automated clustering of home sensor networks to functional regions for the deduction of presence information for medical applications," Wohnen–Pflege–Teilhabe–Besser leben durch Technik , 2014.

[20] S. Blackman, "Multiple hypothesis tracking for multiple target tracking," Aerospace and Electronic Systems Magazine, IEEE, vol. 19, no. 1, 2004, pp. 5–18.

[21] S. Müller and A. Hein, "Multi-target tracking in home sensor networks," in Proc. of the Workshop on AI Problems and Approaches for Intelligent Environments. Prague, Czech Republic: CEUR-WS.org, Aug. 2014, to appear.

[22] D. Reid, "An algorithm for tracking multiple targets," Automatic Control, IEEE Transactions on, vol. 24, no. 6, 1979, pp. 843–854.

[23] D. Cook and M. Schmitter-Edgecombe, "Assessing the quality of activities in a smart environment," Methods of information in medicine, vol. 48, no. 5, 2009, p. 480.