Implementation of a Police Intelligence Analysis Framework

Dafni Stampouli, Matthew Roberts, Gavin Powell, Tomás Sánchez López

Innovation Works, EADS UK
Dafni.Stampouli@eads.com, Matthew.Roberts@eads.com,
Gavin.Powell@eads.com, Tomas.Sanchezlopez@eads.com

Abstract

Police intelligence operations rely on human analysts to search, comprehend and make sense of information during criminal investigations. This is a long and laborious task, in a situation where a quick and effective response is crucial to the prosecution of the offenders. This paper presents the implementation of a Police Intelligence Analysis Framework (PIAF), that provides automated processing of information extracted from witness statements, in order to identify the entities involved in an incident, and provide to the Police operations a complete and accurate situational picture. In order to improve situational awareness and aid the decision making process. This paper outlines the challenges when dealing with intelligence information, highlighting the methods implemented to address those challenges, and provides an overview of a system that is designed to be used in Police post-event analysis. A Police investigation example is provided to demonstrate the use of the system.

Keywords: *Police intelligence, criminal investigations, witness statements*

1. Introduction

During investigation of criminal activities, Police analysts are required to identify the entities involved and their relationships. This intelligence analysis requires the discovery and understanding of information relating to the different entities (such as people, vehicles, locations) involved in the incident under investigation. Therefore, the analyst is trying to comprehend the evidence, determine links among individuals as well as identify leads. A Police Intelligence Analysis Framework is being developed to address the needs of Police intelligence operations [1]. One of the challenges of such operations is that they rely on human-based observations, such as Police reports obtained during investigation, reports from the public (both victims and witnesses), as well as individual informers. Such data by its nature is the result of a subjective assessment, and carries vagueness and uncertainty. Previous work [2-3] analysed in detail the issues with using human intelligence (HUMINT), and presented appropriate methods to combine data in different formats (numeric, textual and ranges). Information fusion methods using the Transferable Belief Model (TBM) were developed to combine vague and incomplete witness statements and use them to identify the described offender from a dataset [3]. This paper presents the implementation of the overall system that uses the developed methods to improve the task of intelligence analysis and provide situation awareness. A simple example of a Police investigation is shown to demonstrate the functionality of the proposed system.

This paper is organised in 5 sections, Section 2 discusses in detail the issues with Police intelligence analysis in dealing with data from human observations, and it presents how entity matching is used to address those. Section 3 introduces situation assessment and the use of entity relationship graphs. Section 4 gives an outline of the proposed implementation and presents an example of Police analysis on how the system can be used to improve situation awareness and aid decision making. Finally, potential future work is outlined in Section 5.

2. Entity Matching

Intelligence operations, and in particular policing investigations, benefit from the use of humans in order to collect, report and interpret information. Humans have the ability detect and report on relationships and inferences which could not be identified by traditional (electronic) sensors, which only measure features and attributes of certain entities. However, the employment of human sensors generates several well reported issues [4]. Human opinions or estimations are prone to errors, biases and imprecisions. When a witness gives a statement trying to describe the offender to the Police, they may estimate the suspect to be 1.77 m when in fact they are 1.72 m. This imprecision is natural and not unexpected. They may give a range of values such as between 1.7 m and 1.80m, or say the suspect was of 'average height', or 'slightly taller than me'. There is need for methods than can process such information and extract knowledge from it. There is also the issue of biases due to personal and cultural characteristics and when more than one witness provide statements, there may be conflicts and disagreements.

Entity matching is a process that combines vague human descriptions of a single entity to provide a more robust description of that entity, and compares that robust fused singular description against a dataset of known entities. An entity can take a number of forms including, people and vehicles. These entities are described using discrete values (i.e. Ford, hatchback, 4 doors) and continuous values and ranges (1.7-1.8m, 60kgs). Often descriptive words can be mapped into a continuous numerical space (blonde hair, tall, overweight) to assist in this entity description. When creating systems that compare descriptions of entities with actual entities it is important to consider the plausibility that the known entity is being described by the vague description. It is easy to ignore the various issues related to data collation that can occur such as vagueness, precision, reliability, confidence, and certainty. These are all important, but subtly different, and all should be taken into account when working with data. One of the key ones when working with human supplied descriptions is that of vagueness of the description. Because of these various effecting factors, it is important that a mechanism is developed that provides a suitable means of comparison between entity attributes that can account for these factors. In addition, attribute comparison and then aggregation of these attributes is required.

Information fusion methods are used to combine entity descriptions to achieve entity matching. Information fusion is the merging of information so that output holds more complete information than the sum of all the inputs. By fusing multiple witness descriptions of a single entity, contextual and subjective discrepancies can be lessened. Using variants of the Transferable Belief Model [5]-[8], which mitigate normal issues related to its usability in real world problems, allows us to fuse together these various descriptions into a single more robust one. This also provides us with an opportunity to look at the descriptions and understand how much conflicting evidence there is, and also identify the most conflicting piece of evidence. It is possible that witnesses are actually looking at, and reporting on, separate entities. Also it is possible that one witness is less of an expert on the subject, or there are other factors affecting the ability to report, such as poor visibility. This could lead to provision of a witness's report that conflicts with the others. This information should be discovered at the outset and removed or accounted for. The system is designed in a way that if a witness gives false information which is conflicting with the rest of the statements, this will become evident at the beginning of the analysis. If a witness statement is removed or ignored, the results are created using the remaining statements.

To demonstrate the fusion process, a simple example of people matching is used. Three witnesses were asked to produce witness statements for an offender. The witnesses were asked to describe offender's height, age, weight, and hair colour. The offender's gender was also noted. These three descriptions were fused together to provide a more robust singular description of the offender, which is then compared with all the people in the dataset. The offender's actual characteristics were inserted to a dataset of 10000 people. The fused description was compared with all the people in the database, and the actual offender came third most probable, with a matching score of 0.7316, with 1 representing a perfect match. This means that in 10000 people there are two other people that match the offender's description very closely (score of 0.8 and 0.76). Several dataset sizes were examined and for datasets of up to 8162 people the offender comes within the top four matches. Further details on the test and evaluation of the fusion methodology can be found in [3].

The same methods can be applied to other entities such as vehicles, and we are currently looking at using *modus operandi* (MO) descriptions to match crimes and criminals and detect re-offenders.

3. Situation Assessment

Situation assessment is the process of estimating the current state of the world environment, by clearly presenting which objects were involved, how they were related, what events took place, and how the situation evolved over time. The aim of the PIAF tool is to assist an analyst in the development of the intelligence picture in order to provide a more timely understanding of the operational environment. One stage further, threat assessment is the process of projecting the understanding of the operational picture into the future, by predicting possible activity which may impact on factors within the incident, and determining the likelihood and the cost associated with the specific events.

Entity relationship graphs (ERG) are currently used to present to the analyst a complete picture of the situation. This allows a quick interpretation of the current situation to be made, in terms of constituent entities and relationships among them. The ERGs are automatically extracted from information gathered about the incident. Possible matches about the entities involved (such as people, location and vehicles) are generated using the entity matching algorithms discussed above. This prompts the analyst for possible suspects, speeding up the analysis process.

An example of the production of the ERG is demonstrated in Figure 1. The incident under investigation is a robbery, and the witnesses have described a person and a run-away vehicle. From the offender's descriptions a list of possible suspects can be created, using the entity matching methods described in Section 2. Each match is ranked according to the matching score, so person 1 matches the description with 0.8 (1 denoting perfect match and 0 no match), persons 2 and 3 match with score of 0.7.

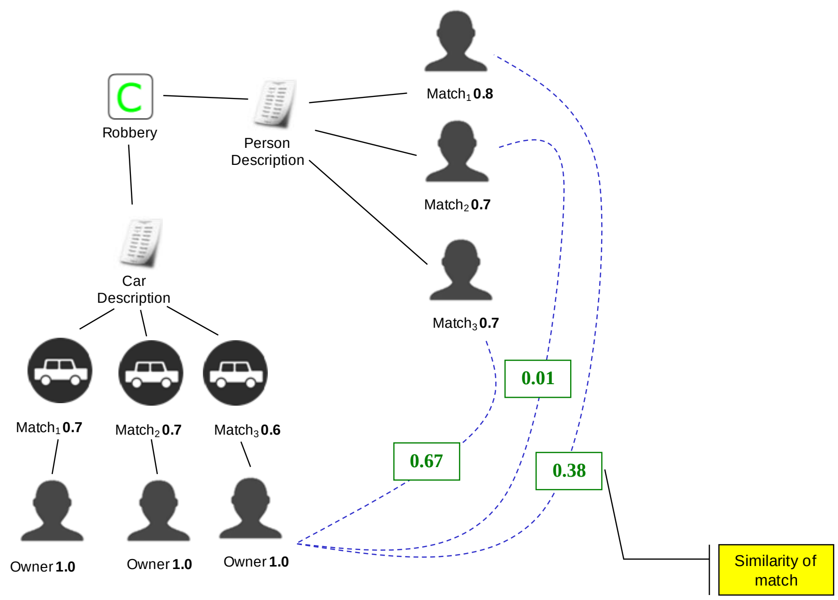


Figure 1. A simple example of an entity relationship graph.

In a similar manner a vehicle description is available and the best matches of the vehicles in the database are presented in the graph, with their associated scores. The next step is to assess whether there is an association between the owners of the cars and the people matching the person description, and that association is given a score. This way the entity graph can grow under the direction and supervision of the analyst. This semi-automated process enables the analyst to remove associations they find irrelevant, and choose the entities they wish to focus on. It also reduces possible human error, as it makes it less probable a good match to be missed, or an important piece of information being ignored, as in a manual analysis. Finally, all the decisions taken both by the system and the analyst can be electronically logged, which may not only help to understand the final results, but would also provide a valuable "trail" for future case audits.

4. System architecture and implementation

The situation assessment tool described in section 3, using the fusion algorithms described in section 2, was designed and implemented as a prototype system. The architecture of the system and its implementation details, as well as a simple example of the use of the tool are described in this section.

4.1 Design and Implementation Technology

The PIAF architecture is designed with flexibility and scalability in mind. Since the system is required to potentially support real-time connectivity to large databases and queries by multiple distributed clients, the architecture must be able to evolve with the increasing needs of both client and server components.

Figure 2 presents a diagram of the PIAF architecture. Essentially, the fusion capabilities of the system are supported at a service layer, which may include one or multiple distributed servers accessible through a network. The server's main functionalities are to:

1. Hold the fusion algorithms that provide the entity matching capabilities
2. Support access to data sources from which the fusion algorithms obtain the information and intelligence to fuse and were fusion results may be stored
3. Provide the architectural components to support the aforementioned functionality and interfaces to access that functionality throughout a network.

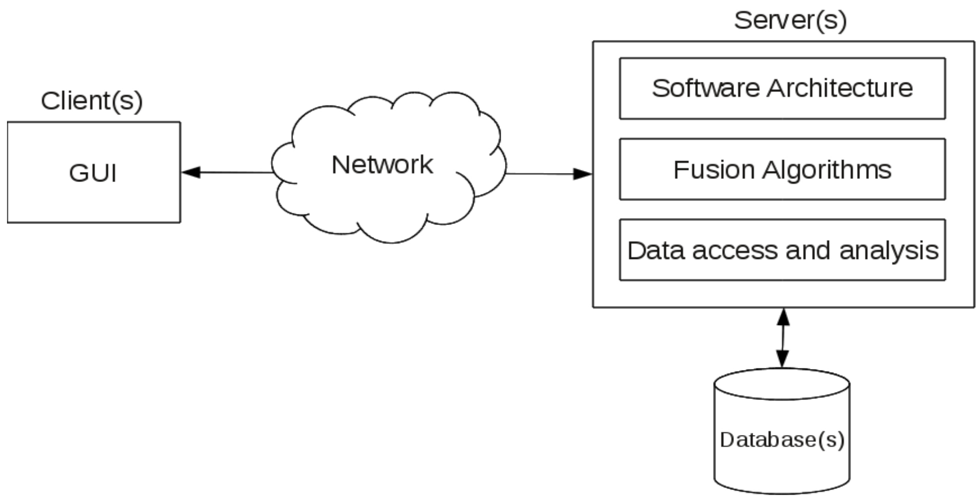


Figure 2. Police Intelligence Analysis Framework Architecture

One or multiple clients would use the server's exposed interfaces to access PIAF's fusion capabilities. A client interacts with the user via a Graphical User Interface (GUI). By means of this GUI, the user has access to all the main systems functionalities while being separated from the software complexity. The design outlined in Figure 2 has been deliberately simplified in order to both convey a clear picture of the architecture and to emphasize the design's benefits. The core of the server component has been implemented using Enterprise Java Beans (EJB), delivering distributed, transactional, secure and portable applications based on Java technology. The Java Beans reside inside an application server, which allows for server side clustering and load balancing. Finally, the client-server communication technology of choice was Web Services, allowing for loosely coupled clients / servers, distribution and location flexibility and independence over the network technology.

4.2 Prototype and Example

A prototype client was implemented using also Java technology. In this section, various screen-shots are used to illustrate an example in which a Police analyst uses the system for the situation awareness of a burglary.

Figure 3 shows a full screen-shot of the client which is divided in 4 main parts:

- Entity Search (Top-right)
- Search results (Bottom-right)
- ERG display (Top-left)
- Entity and relationship information (Bottom-left)

In Figure 3, the analyst initially searches for incidents happening in a specific time period. Note that other entities and other parameters for each entity are also available. Based on a short description of the results, the analyst chooses an incident from the list (for example a burglary), and displays the ERG associated with the incident in the ERG panel. Using the ERG, the analyst gets informed that there are a number of reports, people, addresses and vehicles associated with that burglary. In the specific example, there are two witness reports, each describing a suspect seen at the crime scene. Reports are represented by a document icon with a yellow thunder. When a description is selected (in this example, the one pointed to by a red arrow in Figure 2), the (vague) description of the suspect is displayed on the bottom-left part of the window.

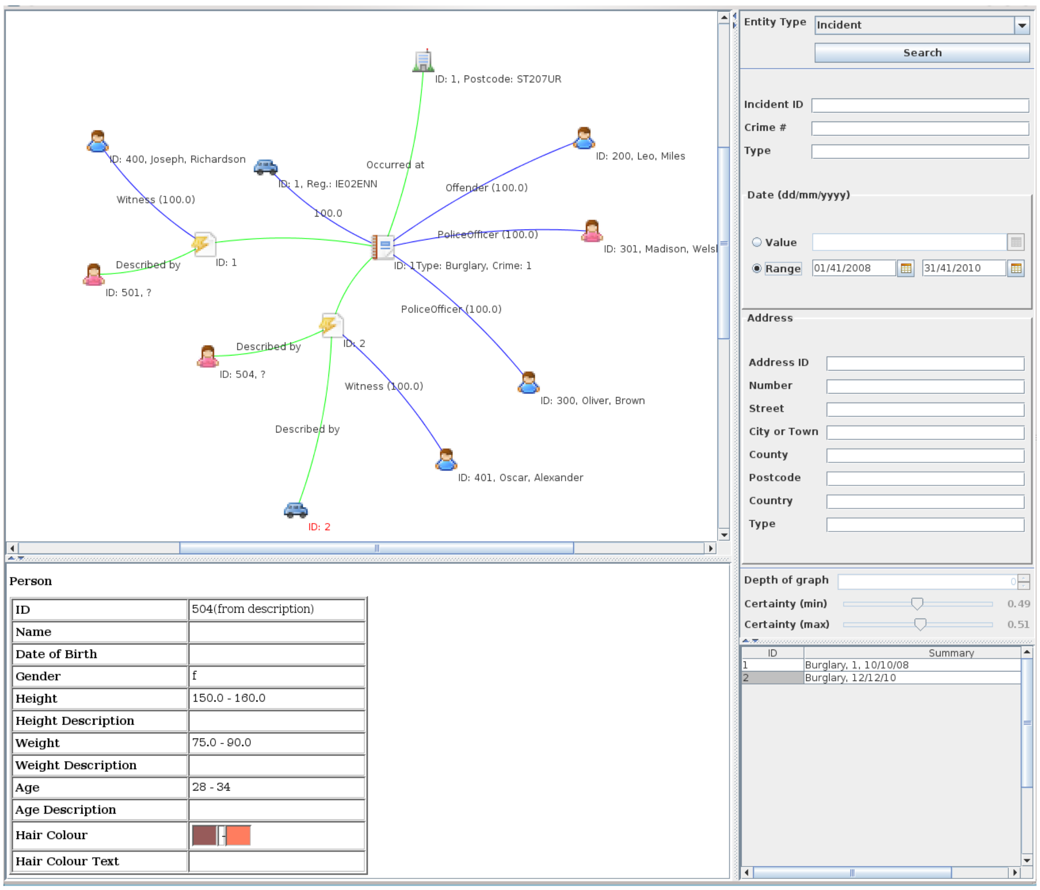
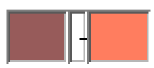
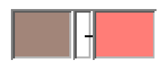


Figure 2. Police intelligence analysis prototype: initial crime search.

The analyst can then investigate whether the selected witness description matches any known person held in the Police database. By clicking the description icon in the ERG, an entity matching process is executed at the server side for all the people in the database. Matches are scored by the fusion algorithms according to their probability, and results are returned to the client and presented in the ERG graph. In Figure 4, only one match, (Grace Thomas), is returned and displayed in the ERG with probability of 69.8%. The analyst is able to confirm this result by matching the second witness description (pointed to by a green arrow in Figure 2). Figure 5 demonstrates that two matches were returned this time, one with very low probability (8.8%) and the other with much higher probability (74.4%). This last match also happens to be Grace Thomas, which corroborates the first witness and confirms to the analysts that Grace Thomas is the suspect that both witnesses saw at the crime scene. Table 1 shows side by side the descriptions of both witnesses and the actual characteristics of Grace Thomas.

Table 1 List of witness descriptions and suspect's actual characteristics.

	Witness Description 1	**Witness Description 2**	**Grace Thomas**
Name			Grace Thomas
Date of Birth			14/07/77
Gender	Female	Female	Female
Height	150-160 cm		155-165 cm
Height Description		Short	
Weight	75-90 Kg		71 – 87 Kg
Weight Description		Overweight	
Perceived Age	28-34	25-32	26-34
Hair Colour			
Hair Colour Text		Strawberry Blond	

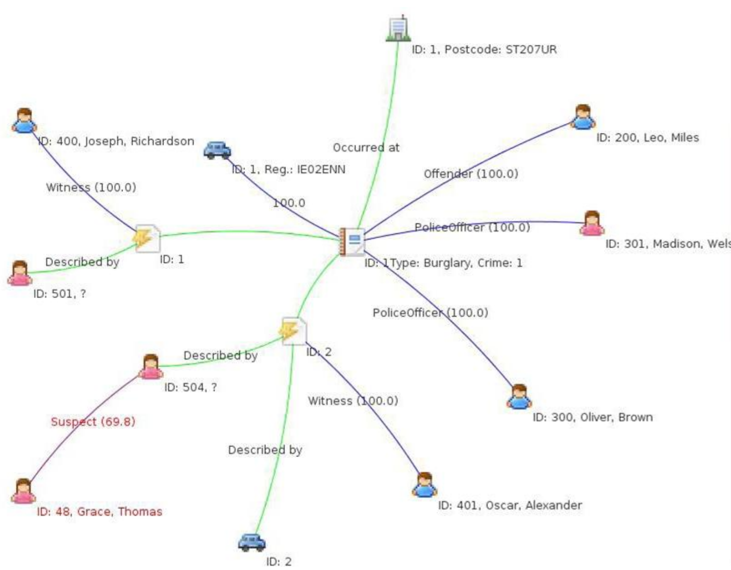


Figure 3. Matching of a witness's description.

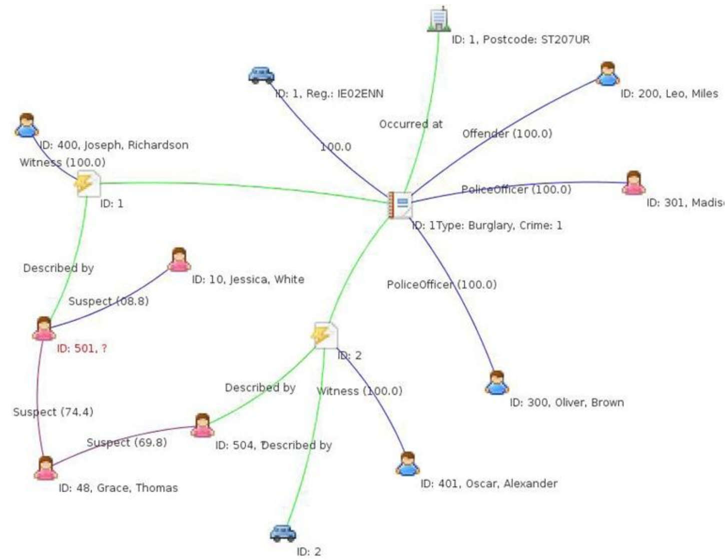


Figure 4. Corroborated witnesses' descriptions.

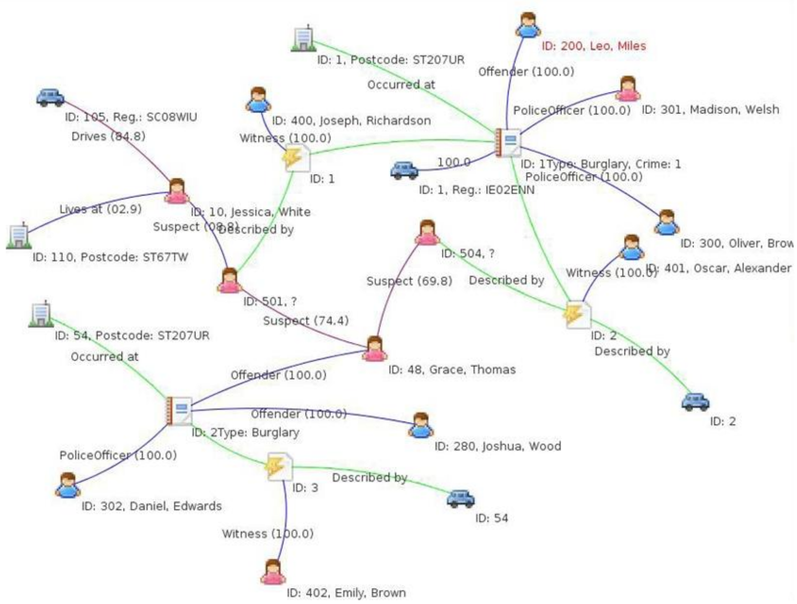


Figure 5. Expanded ERG demonstrates other criminal activities of suspect.

It is also possible to expand the known relationships of various entities in the ERG to investigate links between them. In Figure 6**Figure 5**, the analyst expanded the relationships of Grace Thomas, and discovered that she was in fact convicted for a different burglary. This might prompt a closer investigation of this suspect and give a lead in the crime analysis.

This example demonstrates how information can be made readily available to the police investigation, improving situation awareness and aiding the decision making process. Without the use of such a system, an analyst would be forced to perform a manual search, often across multiple databases, reading lengthy reports and incident records, and trying to identify data relevant to the incident at hand. This manual process is prone to human errors and vital information can be easily missed. This manual process requires days or weeks for an accurate operational picture to be created and the accuracy of the representation depends on the experience and skills of the analyst. Therefore, the proposed system solves the computational burden of the analysis and enables the analyst to drive the investigation without being consumed in the overload of available data.

5. Summary

This paper presents the development and implementation of a software framework, designed to assist the analysis during post-incident Police investigations. The numerous challenges when dealing with vague and incomplete information from human observations are outlined, and the methodologies used to address them are listed. A Police investigation example is presented to demonstrate the use of the system.

Future work involves the extension of entity matching to crimes and criminals based on the descriptions of the crimes and an evaluation of the proposed framework by Police intelligence experts.

Acknowledgements

The authors would like to thank Eleanor Rice, for her guidance and advice relating to Police intelligence analytical theory and practice, and Richard Meeking for his invaluable expertise demonstrated during the implementation of the software.

References

[1] Stampouli, D. Vincen, D. and Powell, G. “Situation assessment for a centralised intelligence fusion framework for emergency services”, 12th International Conference on Information Fusion, Seattle, US, 2009.

[2] Stampouli, D. Brown, M. and Powell, G. “Fusion of soft information using TBM”, 13th International Conference on Information Fusion, Edinburgh, 2010.

[3] Stampouli, M. Roberts, M. and Powell, G. “Who Dunnit? An appraisal of two people matching techniques”, 14th International Conference on Information Fusion, Chicago, 2011.

[4] Hall, D.L.; McNeese, M.; Llinas, J.; Mullen, T. “A framework for dynamic hard/soft fusion”, Information Fusion, 2008 11th International Conference on , Koln, 2008

[5] Shafer, G. “A mathematical theory of evidence”, Princeton University press, 1976, ISBN 0-608-02508-9

[6] Smets, P. and Kruse, “The transferable belief model for belief representation, Uncertainty management in information systems: from needs to solutions”, Kluwer Boston, pp. 343-386, 1997.

[7] Smets, P. “Belief functions on real numbers, International journal of approximate reasoning”, Vol 40, No 3, pp. 181-223, 2005.

[8] Powell, G. Marshall, D. Smets, P. Ristic, B. and Maskell, S. “Joint tracking and classification of airbourne objects using particle filters and the continuous transferable belief model”, 9th International Conference on Information Fusion, 2006.

Authors

Dafni Stampouli received a BEng degree in Electronic Engineering from the University of Central Lancashire, in UK, in 2002. She holds a Masters in data and information fusion from the University of Cardiff, in Wales and a PhD in segmentation of medical images from the University of Central Lancashire. She currently works as a research engineer in Innovation Works in EADS, UK. Her research interest include information fusion, pattern recognition and uncertainty modelling.

Matthew Roberts received a BSc in Computer Science from Cardiff University in 2005. He is currently finishing a PhD at Cardiff University whilst working at EADS Innovation Works. His PhD is sponsored by EADS, and focuses on tracking and classification with wireless sensor networks. Matthew is part of the intelligence fusion group at EADS Innovation Works. His interests focus on belief functions, target tracking, wireless sensor networks, and modelling uncertainty.

Gavin Powell is working within the intelligence fusion group, Innovation Works, EADS where he is within the fusion team, working on a variety of projects which all fall under the umbrella of information fusion. Gavin has focused his attentions on belief functions application, and more specifically the TBM.

Tomás Sánchez López received a BSc and a MSc from the Polytechnic University of Valencia, Spain, in 2004, and a Ph.D in Computer Science in the Korean Advanced Insitute of Science and Tecnology (KAIST), South Korea, in 2008. He joined the Distributed Information and Automation Laboratory, University of Cambridge, where his researched focused in Wireless Sensor Networks, RFID, multi-agent systems and new concepts in distributed systems known as the "Internet of Things". Currently, Tomás works in EADS Innovation Works, and his role involves the supervision and execution of projects related with the distribution and fusion of intelligence.