MOSAIC: A Multi-modal Surveillance System to Enhance Situation Awareness and Decision Making

Richard Adderley[1], Atta Badii[2], Rubén Heras Evangelio[3], Matteo Raffaelli[4], Patrick Seidler[1], and Marco Tiemann[2]

[1] A E Solutions (BI) Ltd., Badsey, UK
{rickadderley,patrickseidler}@a-esolutions.com
[2] University of Reading, Reading, UK
{atta.badii,m.tiemann}@reading.ac.uk
[3] Technische Universität Berlin, Berlin, Germany
heras@mailbox.tu-berlin.de
[4] Synthema srl, Pisa, Italy
matteo.raffaelli@synthema.it

Abstract. With increasing complexity of systems under surveillance, demand grows for automated video-based surveillance systems which are able to support end users in making sense of situational context from the amount of available data and incoming data streams. Traditionally, those systems have been developed based on techniques derived from the fields of image processing and pattern recognition. This paper presents MOSAIC (Multi-Modal Situation Assessment and Analytics Platform), a system which aims at exploiting multi-modal data analysis comprising advanced tools for video analytics, text mining, social network analysis, and decision support in order to provide from a richer context an understanding of behaviour of the system under surveillance and to support police personnel in decision making processes.

Keywords: Multi-modal data mining, social and criminal network analysis; video analytics; semantic interoperability; decision support system.

1 Introduction

As we attempt to monitor increasingly complex behaviour in larger systems, understanding this behaviour from the amount of available data becomes less manageable for the human analyst, possibly leaving a knowledge gap that hinders effective decision-making. One of the key problems for end users of surveillance systems is to 'connect the dots' or to quickly find the few pieces of relevant information from disparate systems and data to establish a greater picture of current situational context. Available information may be structured or un-structured, quantitative and qualitative, of multiple formats, e.g. text, documents, images, videos, or streaming data such as social media or news feeds; be from multiple sources, be of varying quality and reliability, sparse, streaming, and represent rapidly changing situations. Having greater insight would enable officials such as law enforcers, policy makers, and decision

C. Stephanidis (Ed.): HCII 2014 Posters, Part I, CCIS 434, pp. 141–146, 2014.

makers, to deal more effectively with uncertainty, provide timely warning of threats, and to support operational activity by analysing crime.

MOSAIC, a Collaborative Project within the 7th Framework Programme's Research Theme, sets out to improve targeted surveillance by combining data intelligence and advanced video analytics to provide a decision support system for responsible authorities in complex situational contexts. The MOSAIC system enables end users such as police personnel including analysts, administrators, CCTV operators, and staff supervisors to localise and visualise real-time and recorded CCTV video and alerts generated by smart cameras and video analytics nodes to initiate and support police investigations. Providing data mining, text mining and social network analysis, the system also enables the processing, analysis and visualisation of criminal structures and behaviour under investigation. Video analytics and data intelligence tools are being integrated within a unified scalable and modular framework.

The MOSAIC system provides a high-level of usability and effective information display to improve situational awareness and enable faster decision making. Intelligent decision support functionalities support end users in assessing situations and in improving surveillance activities in order to better monitor and respond to identified threats.

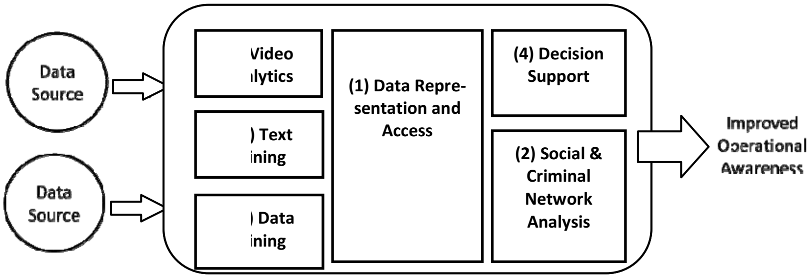


Fig. 1. Overview of the MOSAIC analytical framework with four subsystems: (1) Data representation subsystem; (2) Social & Criminal network analysis subsystem with (2a) Text- & (2b) Data Mining components; (3) Video analytics subsystem; (4) Decision support subsystem.

2 Semantic Interoperability and Data Representation

In order to identify previously unknown connections between disparate data from the enriched operational and analytical picture, the data to be integrated must be queryable using a single query language and return a unified data representation in response to queries. The available information must become semantically interoperable.

Semantic interoperability for MOSAIC involves three main aspects: first, the definition of a semantic domain model, a "world model" which can represent the available information in a way that preserves its meaning; second, the development of a system that organises the available information using the developed model and that also makes it accessible in a unified way; third, the connection to the individual data sources (and to any further "consumers" of the data). The world model for MOSAIC is being defined as an OWL-Lite ontology model [1]. This model represents actors, objects, actions and other relevant information types as subject-predicate-object triples that establish object types, their properties and their relations to other object types. Newly arriving data is analysed for consistency and converted into suitable OWL-Lite instances that are added to the MOSAIC data model stored in the data store. A semantic triple store based on Apache Jena manages the processes of creating, reading, updating and deleting instance data within the semantic representation model. Data in the MOSAIC data model can be queried and updated using the SPARQL query language [2]. The Apache Jena data store implementation has been extended with additional MOSAIC-specific features such as the ability to subscribe with queries in order to allow users to receive notifications when new relevant data has arrived.

The described data representation and data store system allows analysts and operators to query a single data representation for information across information provided by all of the data sources described. The ontology used in MOSAIC extends this by allowing users to make use of the knowledge encoded in the ontology while querying it – a trivial example for this is the ability to query for persons involved in violent crimes without having to enumerate the individual identifiers for violent crimes as might be necessary in a conventional SQL database.

3 Social and Criminal Network Analytics Subsystem

The MOSAIC system will offer data mining support that has been tailored to analysts' need for immediately actionable operational intelligence inside the intelligence cycle. At the lower level, the Entity Resolution Component based on the Apache Lucene framework resolves entities over specified personal attributes, thus alleviating problems regarding poor data quality and supporting identification of all relevant entity data that might not be easily identifiable through global unique identifiers.

Further, following [3], we have formalised analysis tasks using CRISP-DM [4] in conjunction with the intelligence cycle to structure the set of typical analyst tasks. A MOSAIC Data Mining Workbench has been created to assist analysts in manipulating data without the hassle of having to access data from disparate systems. Using the workbench, analysts can search, link, explore, model and visualise data through a process of interconnected nodes; previous knowledge on querying languages is not required. Specific support is provided for: (1) Offender mining and automatic assignment of domain based priorities; (2) Identification of crime series and mapping of known offenders to unsolved crimes; (3) Identification of criminal roles and profiles. Resulting processes are reusable, can be re-run any time taking into account new data, thus accommodate for various possible end user requirements.

In MOSAIC text mining is applied through a pipeline of linguistic and semantic processors (morpho-syntactic tagging, multiword detection and word-sense disambiguation) that share a common knowledge. A domain specific knowledge base with crime patterns, abbreviations and technical terms extracted mainly from anonymised police reports is created. To support analysis tasks specifically, extensive effort is being spent on the recognition of named-entities: dates, addresses, person names, locations, license plate numbers, brands, web entities, bank accounts and phone numbers. Entities are reduced to their semantic roles (agent, predicate, theme, recipient, time and location, i.e. who does what to whom, how, when and where), identified as a result of the dependency parsing. The linguistic processor is then able to extract all kinds of relationships between the entities mentioned above. Entity relationships are visualised in user-friendly network graphs (see Figure 2(b)).

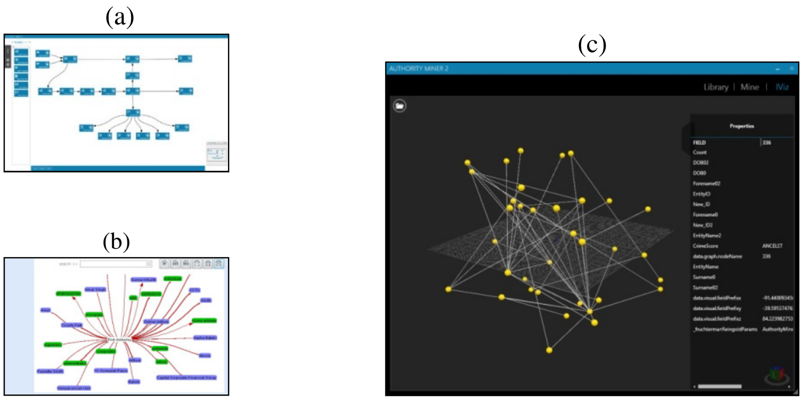
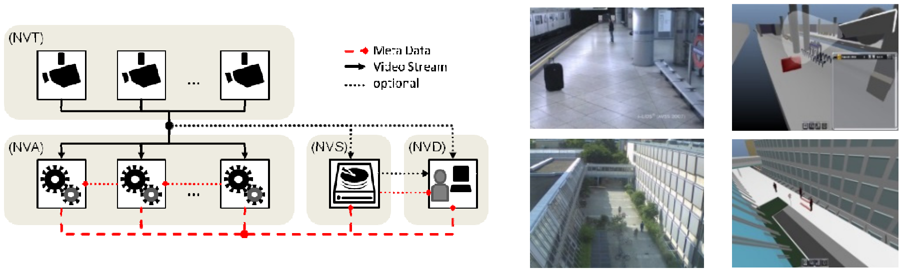


Fig. 2. (a) Data Mining Workbench; (b) Text Mining – Entity Network Visualisation (c) 3D Criminal Network Visualisation and Analysis Tool

Social network analysis specifically is believed to be key in leveraging the operational gap between intelligence analysis and the operational side of businesses [5]. A 3D criminal network visualisation and analysis tool (see Figure 2 (c)) enables the user to conduct Social Network Analysis modified for its application on criminal networks from data accumulated through the data and text mining components, facilitated by the semantically enabled data representation.

4 Video Analytics Subsystem

The video analytics subsystem consists of a number of Networked Video Analytics (NVA) components used to detect events of interest in CCTV footage. Functionalities include visual analytics at low-level such as the detection of change [6] or the computation of optical flow between consecutive video frames [7], at mid-level such as people tracking [8] (also across multiple cameras), and at high level such as crowd behaviour analysis [9], multi-movement identification, automatic detection of human activity [10], mugging detection, etc. The video analytics subsystem is defined as a set of modular NVAs that provide ONVIF [11] metadata in form of events and/or scene descriptions. In this way, mid- and high-level NVAs can use the information gathered by lower levels of analysis and provide higher level semantics (see Figure 3(a)). Combined information provided by the NVAs is collected by the decision support and control sub-system and displayed by means of 3D maps (see Figure 3(b)). This allows for the representation of the gathered information at different levels of abstraction. Furthermore, feedback information of the users can be sent to the NVAs over web services. Following [12], the video analytics sub-system consists of four classes of devices: (1) Network Video Transmitter (NVT) to provide video streams; (2) Network Video Analytics (NVA) for video, audio or metadata analysis; (3) Network Video Display (NVD) for representation of media stream and the gathered information to human operators; (4) Network Video Storage (NVS) for recording streamed video and associated metadata. This architecture provides the common base for a fully interoperable network comprised of products from different network vendors based on Web Services using open and independent standards.



(a) (b)

Fig. 3. (a) ONVIF Network (b) 3D representation of information provided by video analytics. Left: input video. Right: 3D representation

5 Decision Support Subsystem

Decision support in MOSAIC is concerned with the operational support of both intelligence analysts and monitoring systems operators given an environment that is characterised by increasing complexity. Rule-based decision support engines provide configurable support to operators using the MOSAIC system. MOSAIC includes both a high-speed near-real-time rule processing mechanism to provide timely rule processing, and a rule engine that performs complex analyses using the data store's semantic data representation and an advanced rule engine with complex event processing capabilities. Latter can monitor complex events that may be hard to spot by human operators, but that can be defined as sets or sequences of events that taken together either lead to new information or should trigger a specific (re-)action [13].

At the same time, MOSAIC focuses on representing data in context, be it through the presentation of video analytics and event data on a locations' 3D visualisation or the visualisation of relations between relevant actors in an investigation integrating domain priorities and thus producing directly actionable intelligence.

Finally, MOSAIC focuses on empowering users to make better informed decisions. A limited range of direct actions is available through the system, specifically to store information deduced from available data, communicate information to relevant actors and guide further information gathering in particular in the video analytics domain by reorienting cameras and selecting video analytics algorithms based information provided by decision support system components. Additionally, intelligence analysts can simulate intervention strategies based on social network topological measures and domain based priorities for input into operational decision making processes.

6 Conclusions

This paper shows the main building blocks and techniques used in order to realise the vision of the MOSAIC system. Together, the integrated system components aim to empower users to make better use of data that previously has been already available to them but could not be used efficiently and thus also not effectively given personnel and resource pressures affecting organisations such as police forces. The focus of MOSAIC is on improving the operational picture provided to humans, who remain in control and in the loop rather than a complete automation of processes.

Acknowledgments. This work is partially funded by the European Commission (MOSAIC Project, 261776 FP7-SECURITY).

References

1. Horrocks, I.: From SHIQ and RDF to OWL: the making of a web ontology language. Web Semantics: Science Services, and Agents on the World Wide Web 1 (2003)
2. Perez, J., Arenas, M., Gutierrez, C.: Semantics and complexity of SPARQL. ACM Transactions on Database Systems 34(3) (2009)
3. Adderley, R.: Exploring the Differences Between the Cross Industry Process for Data Mining and the National Intelligence Model Using a Self Organising Map Case study. In: Business Intelligence and Performance Management, pp. 91–105. Springer (2013)
4. Shearer, C.: The CRISP-DM model: the new blueprint for data mining. Journal of Data Warehousing 5, 13–22 (2000)
5. Seidler, P., Adderley, R.: Criminal network analysis inside law enforcement agencies – a data mining system approach under the National Intelligence Model. International Journal of Police Science & Management 15(4), 323–337 (2013)
6. Evangelio, R.H., Paetzold, M., Keller, I., Sikora, T.: Adaptively Splitted GMM with Feedback Improvement for the Task of Background Subtraction. Accepted for Publication in: IEEE Transactions on Information Forensics and Security (2014)
7. Senst, T., Geistert, J., Keller, I., Sikora, T.: Robust Local Optical Flow Estimation using Bilinear Equations for Sparse Motion Estimation. In: 20th IEEE International Conference on Image Processing (2013)
8. Eiselein, V., Arp, D., Paetzold, M., Sikora, T.: Real-time Multi-human Tracking using a Probability Hypothesis Density Filter and Multiple Detectors. In: IEEE AVSS (2012)
9. Kuhn, A., Senst, T., Keller, I., Sikora, T., Theisel, H.: A Lagrangian Framework for Video Analytics. In: IEEE Workshop on Multimedia Signal Processing (2012)
10. Acar, E., Senst, T., Kuhn, A., Keller, I., Theisel, H., Albayrak, S., Sikora, T.: Human Action Recognition using Lagrangian Descriptors. In: IEEE MMSP (2012)
11. Open Network Video Interface Forum, http://www.onvif.org
12. Senst, T., Paetzold, M., Evangelio, R.H., Eiselein, V., Keller, I., Sikora, T.: On building decentralized wide-area surveillance networks based on ONVIF. In: IEEE AVSS (2011)
13. Pan, J.: A flexible ontology reasoning architecture for the semantic web. IEEE Transactions on Knowledge and Data Engineering 19(2), 246–260 (2007)