

PAPER SUBMISSION FOR EDF 2014

TITLE OF PRESENTATION:

“VALCRI: Addressing European Needs for Information Exploitation of Large Complex Data in Criminal Intelligence Analysis”

SUMMARY OF THE PRESENTATION

This paper aims to showcase how governments and the public sector are or will be addressing the challenge of Big Data in the context of European law enforcement. We describe a recently approved 18-partner EU-funded Integration Project called VACLRI that will employ the science and technology of Visual Analytics to develop a capability by combining novel visualisation and interaction techniques with powerful analytic software for automated extraction of meaningful information and related text, documents, images and video, and for detecting signatures or patterns across multi-dimensional data that provide early warning or triggers of impending criminal or terrorist action.

TYPE OF PRESENTATION PROPOSED:

Research contribution

CONTRIBUTORS NAMES AND SHORT CV

B.L. William Wong, Leishi Zhang, and Ifan D.H. Shepherd
Middlesex University, London, UK
Contact email: w.wong / i.shepherd / l.x.zhang @mdx.ac.uk

B.L. William WONG is professor of Human-Computer Interaction and head of the Interaction Design Centre. His research interests include the representation design of information to support human decision-making, simulation and training, visual analytics for sense-making, and is inventor of INVISQUE, a novel interactive visual search and query environment. He is or has been Project Coordinator for both the FP7 CRISIS project for the 5-partner UK Visual Analytics Consortium (UKVAC), and a number of other multi-partner projects.

Leishi ZHANG is a Lecturer in visual analytics in the Department of Computer Science at Middlesex University. She received her PhD in Computer Science from Brunel University, UK for her work in time series data analysis and visualization, and then worked as a research associate on a 3-year EPSRC funded project “Visualization with Euler Diagrams”. At Konstanz, she is involved in a number of research projects including DFG funded project “Explorative Analysis and Visualization of Large Information Spaces” and EU funded MOSIPS project as well as various industrial projects. Her research interests include time series data modelling, high-dimensional data projection and interactive visualisation of subspace clusters.

Ifan SHEPHERD is Professor of GeoBusiness at the Middlesex University Business School. His business research interests include personal branding, e-marketing and social media, geodemographics, and the transfer of training and learning. His current research is into the application of videogame technology to geographical information management and analysis, data visualisation, historical reconstruction and training.

VALCRI: Addressing European Needs for Information Exploitation of Large Complex Data in Criminal Intelligence Analysis

B.L. William Wong, Leishi Zhang, and Ifan D.H. Shepherd,
Middlesex University, London, UK
w.wong / i.shepherd / l.x.zhang @mdx.ac.uk

INTRODUCTION

One of the goals of this conference is to showcase how governments and the public sector are or will be making use and benefiting from Open Data, Linked Data and Big Data. In this paper, we hope to show one approach that aims to address some of the challenges law enforcement agencies (LEAs) are facing or will face in the new big data environment. Such an approach is embedded in the VALCRI project – Visual Analytics for sense-making in Criminal Intelligence Analysis – an EU-funded Integration Project (FP7-CP-IP-608142), comprising 18 partners from Europe and the US. It will investigate how visual analytics might be able to support LEAs in developing intelligence-led policing. Such a capability will enhance their ability to be pro-active, rather reactive always waiting for something to happen.

We will describe our plans for how VALCRI will address a number of pressing law enforcement challenges by combining an analyst reasoning workspace comprising novel visualisation and interaction techniques with powerful analytic software for automated extraction of meaningful information and related text, documents, images and video, and for detecting signatures or patterns across multi-dimensional data that provide early warning or triggers of impending criminal or terrorist action. The paper will conclude with a brief discussion about the ethics issues such a technology will present to European society and the rights we value. In the following sections, we will explain what is visual analytics and its origins, criminal intelligence analysis, and discuss a sample of the key science and technology areas of VALCRI.

VISUAL ANALYTICS, INTELLIGENCE ANALYSIS AND BIG DATA

Visual Analytics is the emerging “... science of analytical reasoning facilitated by visual interactive interfaces” (Thomas and Cook, 2004), “... combining automated analysis techniques with interactive visualisations ...” (Keim, et al 2011) that are specially designed to support the interactive dynamics (Heer and Shneiderman, 2012) required to enable real-time analytic interaction with data, “... for effective understanding, reasoning, and decision making on the basis of very large and complex datasets” (Keim et al 2011). Ultimately, the goal of visual analytics is the “... creation of tools and techniques to enable people to: (i) Synthesise information and derive insight from massive, dynamic, ambiguous, and often conflicting data, (ii) Detect the expected and discover the unexpected (iii) Provide timely, defensible, and understandable assessments, and (iv) Communicate these assessment effectively for action.”(Keim et al, 2011). Following the events of 9/11, the U.S. Government set up the National Visualisation and Analytics Centre, NVAC, at the Pacific Northwest National Laboratory, WA, to create the capability for U.S. defence and security agencies to rapidly and interactively making sense of very large, mixed format data sets which may be structured and unstructured, and distributed across many government agencies. This area then came to be known as visual analytics.

What is criminal intelligence analysis? Sometimes known as ‘crime analysis’ (www.interpol.int), it specifically addresses the “... identification of and provision of insight into the relationship between crime data and other potentially relevant data...” (www.interpol.int). This enables officials such as law enforcers, policy makers, and decision

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

The goals of intelligence analysis include: (i) to analyse data and to understand its significance in relation to a problem, and to then reveal the underlying significance of selected target information significance to a specific decision maker (Krizan, 1999). This is not merely reorganizing data and information into a new format; and (ii) to make clear the “sinews of reasoning” (Davis, 1997), i.e. making clear the claims and assumptions, inferences and their bases, and the supporting data and their relationships on which the conclusions are based, especially in complex or controversial issues. The process of transforming data into intelligence is an intellectual endeavour that sorts the “significant from the insignificant, assessing them severally and jointly, and arriving at a conclusion by the exercise of judgment: part induction, part deduction (Millward, in Hinsley and Stripp, 1993) and part abduction” (Moore, 2011, p3).

Big data brings both opportunities and challenges to criminal intelligence analysis. With the expanding use of modern devices such as smart phones, sensors, GPS and video cameras, as well as internet and social media applications such as Facebook and Twitters, more and more data is available for initiating investigations or police actions. This also means advanced systems need to be developed to support rapid linking and processing of large amount of data before it can be turned into actionable intelligence. Such data can be structured or unstructured, static or dynamic, historical or real-time (streaming), from multiple sources (e.g. sensors, CCTV, GPS, news), and of multiple formats (e.g. text, image, video, social media/news feeds).

Over the past few years, an increasing amount of effort has been made towards developing effective visual analytics solutions for handling large complex data and extracting knowledge from it (Zhang et al. 2012) but few of them provide full functionality for handling large data from heterogeneous data sources, in particular, automated or semi-automated event detection based on both structured and unstructured data. Furthermore, solutions for real-time event detection are lacking despite the fact that real-time data plays a more and more important role in intelligence analysis. In terms of data visualization, challenges exist in providing an overview of large amount of interrelated heterogeneous data with temporal and spatial dimensions within a limited display space.

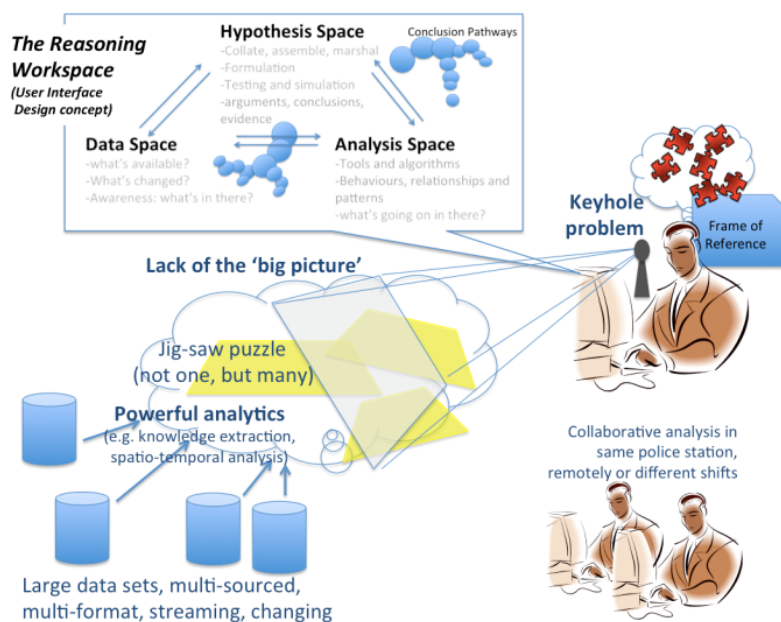


Figure 1. VALCRI operating concept and the Reasoning Workspace (Wong and Varga, 2012)

EXPLOITING BIG DATA IN CRIMINAL INTELLIGENCE ANALYSIS: THE VALCRI APPROACH

The goal of VALCRI is to develop an advanced integrated, multi-function, analyst reasoning workspace that helps analysts in law enforcement agencies to make sense of data by 'connecting the dots' when exploiting and interacting with the very large datasets encountered during criminal intelligence analysis, while ensuring that safeguard are in place to protect against gradual or un-expected erosion of individual liberties. VALCRI will be developed to achieve the following capabilities:

- (i) a Human Issues Framework that combines various human cognition, bias mitigation, social and legal factors into a single principled framework that developers can use to guide and specify the system design. It will identify the ethical, legal and privacy issues that the new technology must address or trade-off (e.g. Burmeister, Duquenoy and Wong, 2014), and it will also examine how technology might hinder human performance in areas such as sense-making and the activation of cognitive biases.
- (ii) an advanced, interactive visualisation-based user interface is guided by the concept of the Reasoning Workspace. It should support the structuring and making visible of arguments and narratives to explain how intelligence has been used; it should enable seamless transition between data exploration, data analysis, and hypothesis formulation; that it will represent uncertainty, and incorporate provenance recording methods to keep track of the human analytical reasoning process.
- (iii) a real-time semantic search and retrieval capability that combines and integrates, and develops where necessary, advanced automated knowledge extraction and thematic clustering based on a self-evolving ontology informed by crime and criminal profiling, from streaming, multimodal data and signature detection. VALCRI will develop automated methods for extracting relevant features to events from historical data. These features will be stored and updated regularly based on new incoming data to help define interestingness of data fragments in real-time. VALCRI will evaluate the interestingness of a data fragment by checking the density of occurrences of relevant features, smoothness of the density curves, the context-coherence of the features and other data-type specific features. The measure will be used to extract potential events from data. Interactive visualizations will be developed at each stage of the data analysis to help users visually explore the data at different abstraction levels. In addition, by applying set theory and piercing decomposition techniques [Stapleton, et al, 2012] complicated overlapping relations between concepts can be decomposed to reduce the complexity of searches such that human analysts can assemble evidence rapidly and reason from a large number of separate text documents.
- (iv) a crime situation re-construction function that is based on spatial-temporal and network technologies for representing important socio-cultural and organizational constructs that are crucial for understanding of the crime and circumstances, and to then project future possibilities
- (v) a secure and scalable distributed processing architecture that is computationally efficient, able to accommodate fast data, and can be delivered incrementally. It must cope with volume, speed, and variety of source and content, within a security architecture supporting the definition and enforcement of security requirements driven by high level privacy and security requirements
- (vi) an anonymised machine deployable dataset that is based on real crimes that is of adequate size and complexity, and to subsequently develop from that process the ability to create and synthesise data that is good as the real data. This data set and process will be made available to the research community to advance research in security.

INTENDED END-USER BENEFITS

Collectively, we anticipate that this research and development programme will lead to the following significant operational user benefits. The intelligence analysts will be able to:

- (i) Quickly locate interesting and suspicious data from across many different data sets – autonomously or with human assistance.
- (ii) Delve deeply into data and to rapidly search across different data sets to retrieve meaningful concepts, and not just key words or frequently occurring words, e.g. a request to “Tell me about Marks and Spencers” would retrieve information about the company, its directors, associated businesses, distribution channels, and other meaningfully associated concepts, rather than occurrences of ‘Marks and Spencers’.
- (iii) Interact with and directly manipulate data and the algorithms in a next-generation information visualisation and dynamic interaction techniques that support a approach that makes visible the reasoning processes and factors considered so as to encourage alternative reasoning strategies, prompts to suggest the need for additional information, the facility to present and discuss alternative points of view, the questioning of assumptions and conditions under which the intelligence information is used, and creation of alternative plausible explanations or lines of inquiry to investigate.
- (iv) Analyse large volumes of data to identify signatures and patterns of triggers and other precursors to criminal activities from low frequency but significant data.
- (v) Use interactive spatial-temporal visualisations for semantically re-constructing criminal events using associated evidence and supporting data, in ways that comply with ethical, privacy and legal expectations.

REFERENCES

- Burmeister, O., Duquenoy, P., & Wong, B. L. W. (2014). Socially responsible design considerations in criminal intelligence analysis. Paper presented at the Ethics, Technology and Organisational Innovation.
- Thomas, J. J., & Cook, K. (Eds.). (2004). *Illuminating the path: A research and development agenda for Visual Analytics*: IEEE CS Press.
- Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (Eds.). (2011). *Mastering the information age: Solving problems with Visual Analytics*. Konstanz: Available for download at <http://www.vismaster.eu/book/>.
- Krizan, L. (1999). *Intelligence Essentials for Everyone*. Washington, D.C.: Joint Military Intelligence College.
- Heer, J., & Shneiderman, B. (2012). Interactive dynamics for Visual Analytics. *Communications of the ACM*, 55(4), 45-54.
- Millward, W. (2001). Life in and out of Hut 3. In F. H. Hinsley & A. Stripp (Eds.), *Codebreakers: The inside story of Bletchley Park* (pp. 17-29). Oxford: Oxford University Press.
- Moore, D. T. (2007). *Critical thinking and intelligence analysis* (2nd ed.). Washington, D.C.: National Defense Intelligence College.
- Wong, B. L. W., & Varga, M. (2012). Blackholes, keyholes and brown worms: challenges in sense making *Proceedings of HFES 2012, the 56th Annual Meeting of the Human Factors and Ergonomics Society, Boston, MA, 22-26 October, 2012* (pp. 287-291). Santa Monica, CA: HFES Press.
- Zhang, L., Stoffel, A., et al., (2012). Visual analytics for the big data era - a comparative review of state-of-the-art commercial systems, In *Proc. of IEEE Symposium on Visual Analytics Science and Technology*, 173-182, 2012
- Stapleton, G., Zhang, L., et al., (2011) Drawing Euler Diagrams with Circles: The Theory of Piercings. *IEEE Trans. on Visualization and Computer Graphics*, 17(7):1020-1032