SAVASA – STANDARDS-BASED APPROACH TO VIDEO ARCHIVE SEARCH AND ANALYSIS

Suzanne Little[2], Kathy Clawson[9], Cem Direkoglu[2], Seán Gaines[10], Francesca Gaudino[1], Roberto Giménez[3], Iveel Jargalsaikhan[2], Emmanouil Kafetzakis[6], Tasos Kourtis[6], Jun Liu[9], Ana Martínez Llorens[7], Anna Mereu[3], Giorgio Montefiore[8], Marcos Nieto[10], Noel E. O'Connor[2], Karina Villarroel Peniza[7], Celso Prados[5], Aitor Rodriguez[4], Pedro Sanchez[4], Alan F. Smeaton[2], Raúl Santos de la Cámara[3], Bryan Scotney[9], Hui Wang[9]*

1. Baker-McKenzie, Italy
2. CLARITY, Dublin City University, Ireland
3. Hi-Iberia, Spain
4. IKUSI, Spain
5. INECO, Spain
6. NCSRD "Demokritos", Greece
7. RENFE, Spain
8. SINTEL, Italy
9. University of Ulster, UK
10. Vicomtech-IK4, Spain

ABSTRACT

This demonstration shows the integration of video analysis and search tools to facilitate the interactive retrieval of video segments depicting specific activities from surveillance footage. The implementation was developed by members of the SAVASA project for participation in the interactive surveillance event detection (SED) task of TRECVid 2012. This year, for the first time, the purpose of the interactive SED task was to evaluate systems' ability to support users in identifying video segments showing an event in a large collection of surveillance videos. Project partners worked together to analyse video and provide a query interface enabling users to search and identify matching video segments. The collaborative integration of components from multiple partners and the participation of end user partners in evaluating the system are the novel aspects of this work.

1. SEARCHING CCTV ARCHIVES

The increasing ubiquity of CCTV and surveillance video systems results in very large archives of footage captured and recorded in remote locations, at different levels of coverage and with different formats, available metadata or searchable indices. Authorised users face many challenges accessing specific footage or finding relevant segments based on semantic descriptions such as 'white car', 'person running' or 'crowded scene'. The key barriers that must be overcome to ensure timely, accurate and legal access to surveillance footage are:

- harmonisation of metadata standards and semantic terms to enable search over multiple archives;
- indexing of video at sufficient granulatity (e.g., person tracking, object detection, time stamping etc.);
- support for integrated and remote access to databases;
- foundational integration of legal, ethical and privacy concepts and
- strong security support.

The SAVASA project (`http://savasa.eu`) aims to develop a standards-based video archive search platform that allows authorised users to query over various remote and non-interoperable video archives of CCTV footage from geographically diverse locations. At the core of the search interface is the application of algorithms for person/object detection and tracking, activity detection and scenario recognition. The project also includes research into interoperable standards for surveillance video, discussion of the legal, ethical and privacy issues and how to effectively leverage cloud computing infrastructures in these applications. Project partners come from a number of different European countries and include technical and research institutions as well as end user, security and legal partners.

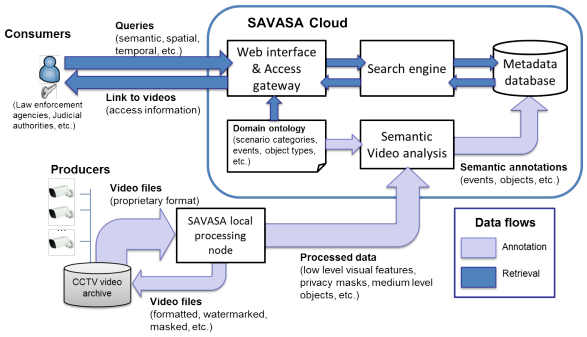

Fig. 1. SAVASA framework

Figure 1 shows the current architecture of the SAVASA platform and illustrates how CCTV footage from "Producers" is analysed in a series of steps to produce a search index using terms from the domain ontology to facilitate advanced search by "Consumers" adhering to the legal and ethical access rights. The project is currently focusing on deploying components within the cloud infrastructure and ensuring security and access enforcement. The project has also conducted a survey of standards relating to surveillance video and is evaluating video transcoding options to identify the best formats to use to support user's requirements. A key contribution of the project is the development of an ontology organising semantic terms relating to people, events and objects to describe scenarios that are relevant for our end users. This ontology will be exploited to improve scenario recognition and semantic annotation.

Contact author: suzanne.little@dcu.ie

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 285621, project titled SAVASA.

The novelty of the SAVASA project lies in the active participation of end users to guide and determine the legal, ethical and security requirements when building a cloud-based integrated search system for semantically annotated surveillance videos.

2. SEMANTIC ANNOTATION OF VIDEO

To facilitate the aims of the SAVASA project, we took part in the TRECVid Interactive Surveillance Event Detection task 2012 [1]. TRECVid is an annual benchmarking exercise sponsored by the US National Institute of Standards and Technology (NIST) with the aim of stimulating video information retrieval research and improving the performance of systems using large, challenging, realistic and noisy datasets for real world problems. Surveillance Event Detection using CCTV footage has been a TRECVid task for the previous five years but, due in part to the lack of significant improvements in detection rates, was changed this year to include an interactive element. Previously, a set of test (unannotated) videos would be processed by one or more event classifiers and the ordered list of possible matches would be evaluated to determine the system's performance. This year a user interface could be used to identify video segments in the test set with a 25 minute search time limit per user per event class.

Our approach was to combine individual methods for video analysis and annotation and provide a dashboard style search interface (Figure 2) that enabled the user to view results for various algorithms and filter them by factors such as confidence, level of motion, camera, number of people etc. The interface was translated into Spanish to support user partners from Vicomtech-IK4, IKUSI, RENFE and HIB participating in the evaluations. The semantic video analysis and annotation tools performed a range of functions including person tracking, region of interest mapping, event recognition using region-based identification and sparse trajectories and varying combinations of features for each approach. The outputs from these systems were converted to a standard description format (ViPER [2]) and uploaded to a database for searching through the user interface. Results were ranked using normalised confidence values provided by each system.

The challenges we faced in integrating different analysis and classification techniques included choosing suitable formats to exchange descriptors and upload resulting annotations, normalising the confidence values to merge results' lists and choosing a fusion method to build the final list of results for submission from the list of segments found by all users. The results of our evaluation were competitive within the TRECVid framework but still show very low performance for any practical application purpose and provided us with interesting new directions to follow. The feedback given by the users regarding the interface, search options and their priorities in surveillance video search was extremely valuable. Details regarding the classifiers and their evaluation performance can be found in [3, 4].

3. DEMONSTRATION

This demonstration will show the evaluation interface used in the SED task by the end user partners from the SAVASA project. It illustrates how video analysis techniques from independent sources can be brought together to support interactive identification of surveillance events. Through the interface a specific event is chosen and by selecting different classifiers (systems) a ranked grid of animated GIFs shows the segments annotated with the event (Figure 2). The user can browse results and choose matching video segments to discover as many events as they can in the 25 minute time limit.

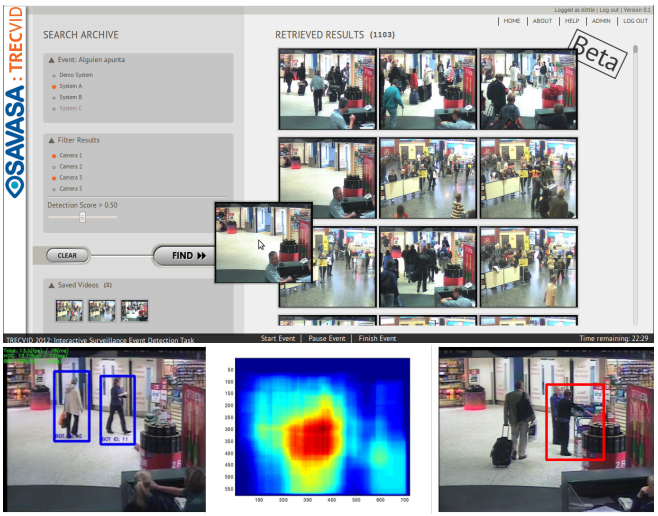

Fig. 2. Screenshot of SAVASA interface for the TRECVid 2012 interactive surveillance event detection task (top) and Classifer examples: tracking, ROI heatmap, Pointing (bottom)

The TRECVid dataset, comprising CCTV footage from an airport, will be used to populate the demonstration prototype. This is a complex real-world dataset with difficult to identify activities, multiple cameras, a range of scales and significant variations in crowding and occlusions. A secondary contribution of this demonstration is the opportunity to explore and discuss the complexities of real-world activity recognition in surveillance video by choosing to apply different systems and confidence value filtering to change the results' list. A screen capture of the interface in action is available at `http://youtu.be/ybJyHWRgJBc`.

4. REFERENCES

[1] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, B. Shaw, W. Kraaij, A. F. Smeaton, and G. Quénot, "TRECVid 2012 – An Overview of the Goals, Tasks, Data, Evaluation Mechanisms and Metrics," in *Proceedings of TRECVID 2012*. NIST, USA, 2012.

[2] D. Doermann and D. Mihalcik, "Tools and techniques for video performance evaluation," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 4, pp. 167–170 vol.4.

[3] S. Little, I. Jargalsaikhan, C. Direkoglu, N. E. O'Connor, A. F. Smeaton, K. Clawson, H. Li, M. Nieto, A. Rodriguez, P. Sanchez, K. Villarroel Peniza, A. Martínez Llorens, R. Giménez, R. Santos de la Cámara, and A. Mereu, "SAVASA Project @ TRECVid 2012: Interactive Surveillance Event Detection," in *TRECVid Workshop*, 2012.

[4] S. Little, I. Jargalsaikhan, K. Clawson, H. Li, M. Nieto, C. Direkoglu, N. E. O'Connor, A. F. Smeaton, B. Scotney, H. Wang, and J. Liu, "An Information Retrieval Approach to Identifying Infrequent Events in Surveillance Video," in *ACM International Conference on Multimedia Retrieval*, 2013, to appear.