Interactive Surveillance Event Detection at TRECVid2012

Suzanne Little*
Iveel Jargalsaikhan
CLARITY, Centre for Sensor Web Technologies
Dublin City University, Ireland

Kathy Clawson
Hao Li
University of Ulster
Belfast, United Kingdom

Marcos Nieto
Vicomtech-IK4
San Sebastian, 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 that depict a specific activity (event) in a large collection of surveillance video footage. 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.

Categories and Subject Descriptors

I.2.10 [**Vision and Scene Understanding**]: Video analysis; H.5.2 [**User Interfaces**]: Benchmarking

Keywords

TRECVid, surveillance video, event detection

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'.

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 (Figure 1). 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.

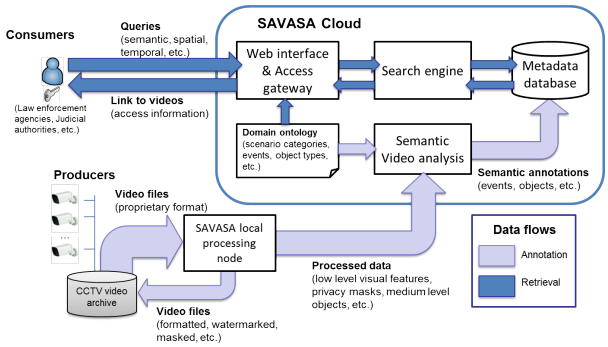

Figure 1: SAVASA framework (v1.0)

2. TRECVID SED INTEGRATION

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.

*Contact author: suzanne.little@dcu.ie

ICMR'13, April 16–20, 2013, Dallas, Texas, USA.

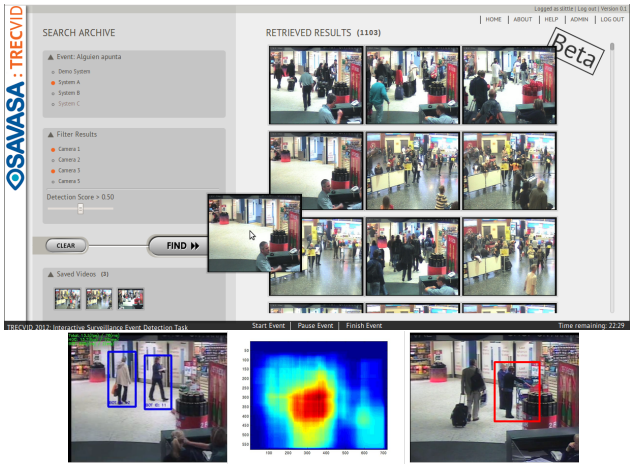

Figure 2: Screenshot of search interface (top) and Classifer examples: tracking, ROI heatmap, Pointing

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.

Two main methods were used to identify video segments showing one of three possible events – ObjectPut, PersonRuns, Pointing. The first classified events based on descriptors from motion trajectories. The trajectories were calculated using salient points identified by Harris Corner detectors and tracked using the Kanade-Lucas-Tomasi (KLT) algorithm. Trajectory length was empirically set to 15 frames and the trajectories described using HOG, HOF, MBH and TD descriptors. Descriptors were clustered using a Bag-of-Features approach to reduce the dimensionality before being classified via a trained SVM.

The second classifier looked at region-based identification and used two configurations to compare different approaches – Optical Flow features with a Hidden Markov Model classifier and dense SIFT features using Bag-of-Features and an SVM classifier. A method for person detection and tracking was implemented to provide input and to generate metrics about person density and activity to use in the interface. Persons were detected using HOG descriptors and tracked via a Rao-Blackwellized Data Association Particle Filter previously shown to produce good multiple object tracking results even with sparse detections. Due to the crowded nature of the scenes, the performance of the person detection was insufficient to be fully integrated in this stage and fixed regions were used instead.

The person tracking provided input for manual region of interest identification to determine *a priori* probability of an activity occuring in a region and confidence values of the region-based classifier were adjusted accordingly.

The challenges we faced in integrating difficult 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 [2].

3. PROPOSED 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 25min time limit.

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`.

Acknowledgements

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. Thanks to Kevin McGuinness for providing code for some interface components from the AXES project (`http://www.axes-project.eu`) search interface.

4. ADDITIONAL AUTHORS

Cem Direkoglu (DCU, Ireland), Noel E. O'Connor (DCU, Ireland), Alan F. Smeaton (DCU, Ireland), Jun Liu (UU, UK), Bryan Scotney (UU, UK), Hui Wang (UU, UK), Seán Gaines (Vicomtech-IK4), Aitor Rodriguez (IKUSI, Spain), Pedro Sanchez (IKUSI, Spain), Ana Martínez Llorens (RENFE, Spain), Karina Villarroel Peniza (RENFE, Spain), Roberto Giménez (HIB, Spain), Raúl Santos de la Cámara (HIB, Spain), Anna Mereu (HIB, Spain), Celso Prados (INECO, Spain), Emmanouil Kafetzakis (NCSR "Demokritos", Greece)

5. REFERENCES

[1] P. Over, G. Awad, *et al.* TRECVid 2012 – An Overview of the Goals, Tasks, Data, Evaluation Mechanisms and Metrics. In *Proceedings of TRECVID 2012*. NIST, USA, 2012.

[2] S. Little, I. Jargalsaikhan, *et al.* SAVASA Project @ TRECVid 2012: Interactive surveillance event detection. In *TRECVid 2012 - TREC Video Retrieval Evaluation Workshop*, Gaithersburg, MD, 2012.