Unsupervised Activity Analysis and Monitoring Algorithms for Effective Surveillance Systems

J.-M. Odobez[1], C. Carincotte[2], R. Emonet[1], E. Jouneau[2], S. Zaidenberg[3], B. Ravera[4], F. Bremond[3], A. Grifoni[5]

[1]Idiap Research Institute, [2]Multitel, [3]INRIA, [4]Thales Communication, [5]Thales Italia

Abstract. In this demonstration, we will show the different modules related to the automatic surveillance prototype developed in the context of the EU VANAHEIM project. Several components will be demonstrated on real data from the Torino metro. First, different unsupervised activity modeling algorithms that capture recurrent activities from long recordings will be illustrated. A contrario, they provide unusuallness measures that can be used to select the most interesting streams to be displayed in control rooms. Second, different scene analysis algorithms will be demonstrated, ranging from left-luggage detection to the automatic identification of groups and their tracking. Third, a set of situationnal reporting methods (flow and count monitoring in escalators and at platforms as well as human presence at lift) that provide a global view of the activity in the metro station and are displayed on maps or along with analyzed video streams. Finally, an offline activity discovery tool based on long term recordings. All algorithms are integrated into a Video Management Solution using an innovative VideoWall module that will be demonstrated as well.

1 Introduction

The demonstration will show different modules developed in the EU VANAHEIM project[1] and that have been exhibited in the Turin Automated Metros security room to participants of the EXPO Ferroviaria fair that was held in Torino at the end of March[2]). The modules feature different innovative surveillance algorithms developed within the project that can increase the effectiveness of existing public transport surveillance systems. The modules are summarized below.

2 Activity modeling for stream selection

There is an ever increasing number of sensors deployed in the real world, being it in large scale sensor networks or closed-circuit television (CCTV) networks. Exploitation of these cameras fulfil several needs, safety, security, or efficiency,

---

[1] http://www.vanaheim-project.eu
[2] http://www.expoferroviaria.com/eng/page.cfm/link=57

e.g., through the identification of typical flux and detection of congestion that can be communicated appropriately to users.

However, most of the time, surveillance network videos are never watched, simply due to the odds: a control room may have 28 monitors to supervise more than 1100 cameras as in our case, and thus the probability to watch the right streams at the right time is therefore very limited. Thus, along with top-down user-based protocols for visualizing and browsing the video networks and the detection of predefined event, bottom-up automatic and content-based selection systems that provide surveillance operators with the most salient and informative data streams could can provide a useful tool to address the selection of streams to be displayed and watched.

The proposed demonstration will address this issue, and show the results of recently published algorithms on unsupervised content-base stream activity modeling and abnormality level estimation [1–4].

Audio stream analysis. Audio is a complementary modality that can easily be coupled with video for surveillance and scene understanding. This topic was addressed in the VANAHEIM project for the detection of abnormal audio events. In public transportation context, the acoustic environment is complex, and can be viewed as a superposition of many single audio events that are considered as normal (people discussing, trains arrivals and departures, silences, etc.). In this view, normal audio ambiance models are estimated in an unsupervised way using models such as Gaussian Mixture Models or One Class Support Vector Machines [5]. The demonstrator will also illustrate these methods with several kinds of abnormal events mixed with real audio ambiances.

3   Scene analysis algorithms

In a second step, different scene analysis algorithms will be demonstrated. These will range from multi-object tracking [6] the extraction of behavioral cue like body and head pose [7] through coupled adaptive learning, that will be illustrated on real video from offline processing, to real-time processing algorithms like those described below.

Static-luggage detection. The algorithm relies on a multi-layer background subtraction method [8] by distinguishing recent background layers from old and long term background layers and foreground regions. It is enhanced by a human-detection algorithm [9] to remove false alarms due to people waiting and hence remaining static for quite some time.

Group detection and behavior analysis. The demonstration will show real-time group detection and tracking in metro data from various views and detected events corresponding to predefined behaviors of interest. The group detection and tracking method works by first tracking mobile objects and then grouping them recursively over time maintaining a spatial and temporal group coherence, using proximity as well as walking similarity (based on speed and direction) [10].

Group behavior is analyzed according to predefined formal scenarios operating on properties of tracked groups [11] .

Flow monitoring. Several crowd characterization approaches will be shown based on crowd density estimation through interest point extraction and tracking or occupancy rate estimation.

4 Situational reporting

Finally, a full set of methods that provide estimates of people locations and numbers according to the different camera settings: people flow counting in escalators; people density at platforms and human presence at lift achieved using straight human detection methods [12], and, whenever possible, multi-object tracking relying on Probabilistic Occupancy Maps (POM). All the computed information is back-projected in real time on the infrastructure map of Turin metro or on individual scene views, providing the operators with synthetic views of the metro occupancy and activity.

5 People Activity Discovery Tool

Finally, an offline line tool that aims at analyzing long term recording will be demonstrated. It aims at three different tasks summarized below: i) learning of (floor) activity zones within the scene; ii) learning of activity classes, and iii) calculation of activity statistics. The learning of these activities is done online [13].

The Zone Discovery task consists in identifying the scene activity zones corresponding to entry/exit zones, waiting zones and zones where people interact with the station equipment (i.e. vending machine areas). The processing is done on-line, analyzing video chunks of one hour as they become available. Roughly explained the process is as follows. Stopping points from mobile objects are detected by trajectory analysis and clustered to find a first set of activity zones. The zones that overlap are merged together to get one single activity zone, and activity zones are updated on-line with the analysis of long term data.

Learning of activity classes is a higher level analysis process to extract the main activity patterns of people in the station. It works by linking low-level tracking to the previous learned zones. The whole activity observed from the scene can then be reported following the behaviors (mobiles inside a zone, or moving from one zone to another one) inferred from the learned zones, and build activities that are the combination of all zone-based inferred behaviors.

Based on activity zones and activity patterns, descriptive statistics can be calculated such as the mean occupancy of a zone, the mean time spent in a zone and how zone occupancy changes over time. Similarly, the most common (frequent) activity patterns can be extracted as well as rare activities.

6   Demonstrator

The demonstrator consists of a professional Video Management Solution developed by Thales Italia, where all the analytics modules developed for the VANAHEIM project have been integrated. The user of the system can easily launch any available algorithm, select the audio or video stream to use, and configure the graphical overlay (videowall) where to show the analytics' results. Given that no live feed can be provided, algorithms will be applied (with online processing) on 26 real data streams pre-recorded from one of the metro stations in Torino.

References

1. Emonet, R., Varadarajan, J., Odobez, J.M.: Extracting and locating temporal motifs in video scenes using a hierarchical non parametric bayesian model. In: CVPR. (2011)
2. Emonet, R., Varadarajan, J., Odobez, J.M.: Multi-camera Open Space Human Activity Discovery for Anomaly Detection. In: IEEE International Conference on Advanced Video and Signal-Based Surveillancei (AVSS). (2011)
3. Jouneau, E., Carincotte, C.: Particle-based tracking model for automatic anomaly detection. In: International Conference on Image Processing. (2011)
4. Jouneau, E., Carincotte, C.: Mono versus multi-view tracking-based model for automatic scene activity modeling and anomaly detection. In: EEE Int. Conf. on Advanced Video and Signal-Based Surveillance. AVSS. (2011)
5. Lecomte, S., lengell, R., Richard, C., Capman, F., Ravera, B.: Abnormal events detection using unsupervised one-class svm - application to audio surveillance and evaluation. In: IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance. (2011)
6. Heili, A., Chen, C., Odobez, J.: Detection-based multi-human tracking using a crf model. In: ICCV Workshop Visual Surveillance. (2011)
7. Chen, C., Odobez, J.M.: We are not contortionists: coupled adaptive learning for head and body orientation estimation in surveillance video. In: CVPR. (2012)
8. Yao, J., Odobez, J.M.: Multi-layer background subtraction based on color and texture. In: IEEE International Conference on Computer Vision and Pattern Recognition. (June 2007) 1–8
9. Yao, J., Odobez, J.M.: Fast human detection from joint appearance and foreground feature subset covariances. Computer Vision and Image Understanding, $\mathbf{115}(10)$ (2011) 1414–1426,
10. Zaidenberg, S., Boulay, B., Garate, C., Chau, D., Corvee, E., Bremond, F.: Group interaction and group tracking for video-surveillance in underground railway stations. In: Int. Workshop on Behaviour Analysis, ICVS. (2011)
11. Zaidenberg, S., Boulay, B., Brmond, F.: A generic framework for video understanding applied to group behavior recognition. In: IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS). (2012)
12. Descamps, A., Carincotte, C., Gosselin, B.: Person detection for indoor video-surveillance using spatio-temporal integral features. In: Interactive Human Behavior Analysis in Open or Public Spaces Workshop (INTERHUB). (2011)
13. Patino, L., Bremond, F., Thonnat, M.: Online learning of activities from video. In: IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS). (2012)