On-Line Object Behaviour Analysis for Surveillance Systems

Vítězslav Beran, Roman Juránek, Jozef Mlích, Pavel Žák, Adam Herout, Pavel Zemčík
Brno University of Technology
Faculty of Information Technology
Department of Computer Graphics and Multimedia
Božetěchova 2, 612 66 Brno, CZ
{beranv,ijuranek,imlich,izakpa,herout,zemcik }@fit.vutbr.cz

Abstract

Automatic surveillance systems are an important emerging application of on-line simple or compound event detection algorithms in video or audio data. The nature of such systems implies several requirements on the used algorithms. The system ability to give the response on-line is the main topic addressed in this work. The paper predefines the requirements for each system recognition module that must work in real-time or faster.

This paper describes and analyses several event recognition algorithms based on video processing, concerning the applicability in on-line surveillance systems. Two main trends are addressed with focus on the methods' speed and robustness. Statistical approaches are the basis for simple event detectors whereas spatio-temporal rules define compound event detectors. Examples of statistical approaches are ad-hoc bicycle detector, dog detector based on AdaBoost classifier and exploitation of Hidden-Markov-Models for trajectory classification. The compound event approach is demonstrated on detection of "dangerous occurrence of the person on the platform edge" in underground scenario. The methods process the video data from standard low-resolution CCTV surveillance system.

The developed approaches are evaluated on real data and applied in the real sites in underground scenarios.

1. Introduction

The surveillance systems monitor the behavior of people or objects. One of several types of surveillance is computer surveillance and more specifically computer surveillance based on camera devices. The present systems are mostly used to record collections from a network of sensors (cameras, microphones, etc.). The multimedia streams might be delivered to the control room and watched on-line. Later on, the recorded multimedia collections can be explored and processed off-line. The recording systems widely use some approaches for recognizing the type of the activity, e.g. signals change in the time domain or differs from some predefined pattern (foreground object detections, etc.). The activity recognizers improve the compression ratio when recording multimedia streams. Such recognizers work in real-time and give the result on-line, but they provide more-or-less binary output (activity / no-activity).

The data mining techniques promise automatic extraction of relevant semantic metadata from raw multimedia. Extracted structured knowledge could potentially represent a useful source of information if stored and automatically analyzed [1]. But the process and results are off-line.

The current surveillance systems working in real-time are predominantly designed to recognize predefined specific events [2], such as vehicle, people, vehicle's license plate, etc. and often contain accelerated modules using hardware (GPU, DSP+FPGA, etc.). Hardware acceleration for recognition is promising [3], but is difficult to design and modify the modules so far.

The on-line response in surveillance systems is crucial when the circumstances require immediate action. One example could be detection of the "dangerous occurrence of the person on the platform edge" at the underground or railway stations. The personnel in the control room need an automatic support coming from the surveillance system warning him in case of a potentially dangerous situation.

This paper describes the progress made towards the development of on-line event recognition algorithms. Two main trends have been investigated. The first one relies on event detectors learned using statistical models; they require a training dataset of reasonable size at some point. Examples of statistical approaches are the detection of objects (dog, bicycle) which can provide alarm events about unwanted objects (e.g. big dogs). Another example is the analysis of trajectories of people based on Hidden Markov Models (HMM).

The second type of event recognition approaches relies on event defined through a set of spatio-temporal rules involving the components defined in the ontology. Such compound event approach is demonstrated on detection of "dangerous occurrence of the person on the platform edge" on the underground or railway stations.

2. Detection of Bicycles

For the algorithm of bicycle detection, we have developed and constructed a specialized detector composed of several standard image processing and object-detection techniques combined together ad hoc. The method is described in details in [4].

Our method detects wheel-like shapes. Analysis of the actual data from the scenario (subway access corridors, platforms) showed that it would be impossible to look for the whole shape of a bicycle, because the bicycle can often be partially covered by a human figure. On the other hand, the shape of a wheel of reasonable dimensions does not appear in the videos just except for the case of bicycles. The analysis of the data also allows us to use a prior knowledge about the supposed color of the wheels.

The changes in the bias value in the wheel color model results in several distance maps (layers), when each layer is further processed separately. Such approach improves the robustness to lighting changes. The Figure 1 shows the original example and three increasing layers.

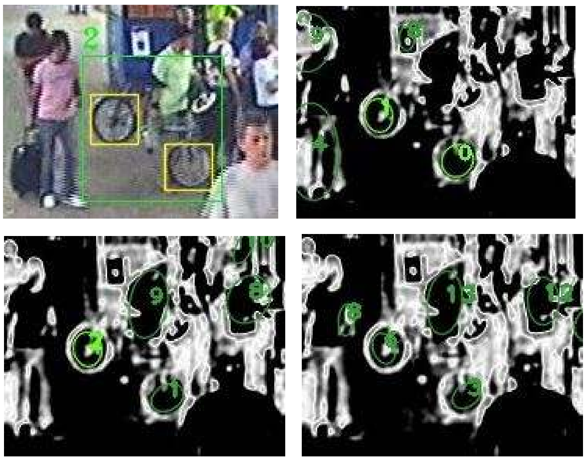

Figure 1: Original image example and corresponding distance maps.

On each layer the wheel candidates are detected as the contours of moderate size and curvature. Each candidate is described by several simple features. Two most interesting simple features are based on multi-scale template matching and ellipse profiling [4].

The detector was trained and evaluated on images from video of Roma undergrounds. Still images containing bicycles were taken and manually annotated that gives wheel samples. The ROC curve on the figure below was generated by processing the wheel detector on each image and the wheel candidates were compared to the ground truth data. The ROC curve describes the trade-off between true positive and false positive detection rates. The gray curves are 95% confidence interval of the fitted ROC curve. The detector performance is good for real applications, when one can expect the object (wheel) in more than one image, so even keeping the false positive rate low gives reasonably high probability of true positive detection.

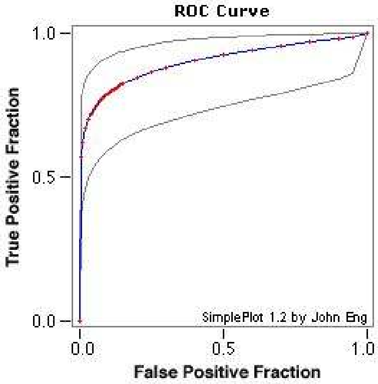

Figure 2: ROC curve of the wheel detector.

The bicycle detector is sensitive to wheel-like objects that do not necessary need to be a bicycle. False positives might be baby carriages, logos on clothes or shiny bald spots. High-level classification such as compound event classifier can be designed to cope with this type of errors.

3. Detection of Dogs

One of the general 2D classification method was used for the dog detection task. The developed recognizer is based on boosting techniques in combination with Haar features and newly developed LRD features [5]. For the training, WaldBoost algorithm [6] was used, which is a modification of well know AdaBoost approach. The method is described in details in [7].

The samples that are necessary for the training are small images of object we want to detect. The specific problem connected with detecting dogs is that the variety of dog shapes (different orientation, posture, etc.) is very large and also texture and brightness of dogs varies in wide range. Unlike, for instance, in the case face detection, the dog detection lacks visually well defined class. Single WaldBoost classifier is, therefore, not able to detect dogs viewed from arbitrary angle. For the above reason, we decided to train classifier of dogs viewed from the profile only. While this limitation may seem relatively severe, it does not introduce any serious limitation from the application point of view as the objects in the video sequence can be tracked and whenever the classifier detects a dog (seen from the profile, the whole track is known to represent the dog.

Samples for the training were collected from publicly available images on the Internet. The images were annotated by hand, divided to training and testing sets and used as input for the WaldBoost training.

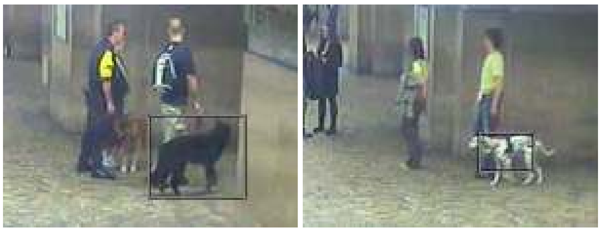

Figure 3: On-line dog detection.

The collected images are from different sources and thus the quality varies. Also, the conditions (background, lighting) do not correspond to the conditions in the target environment (which is underground station). This presumably leads to worse performance of the classification. The dataset now contains hundreds of dog images which is acceptable for WaldBoost training.

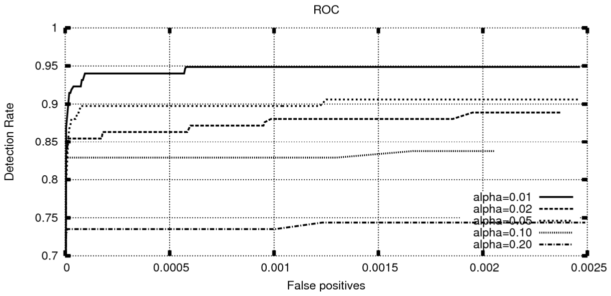

Figure 4: ROC curves for different settings of the training algorithm.

The trained classifiers are then used to detect dogs in input image. The image is scanned by floating window, evaluating the classifier on each position in different sizes. Positive responses of the classifier are passed to non-maxima suppression to avoid multiple detection of a single object.

The experiments show that WaldBoost based classifiers are suitable for dog detection that is performed under constrained conditions. Precision of presented classifier can be increased by expanding the training dataset. Classification performance can be further improved by using motion-sensitive classifiers.

4. People behavior

The people behavior recognizer was constructed using object tracking methods and Hidden Markov Model pattern recognition approach. More detailed description of the approach itself is in [8].

On the contrary with previously mentioned tasks, this approach works with spatio-temporal information. Basic tracking methods were exploited optimized to work in real-time. Due to strong accuracy dependence of the classifier to the tracker, we pre-process the trajectories to remove strong noise.

The people behavior recognition in general is very complex task, maybe even impossible task. Our approach is based on assumption that patterns in person movement corresponds with certain subset of person behavior relevant for surveillance task. For selected camera (e.g. scenario), three main people paths were defined, which represents the main behavior classes (see in Figure 5).

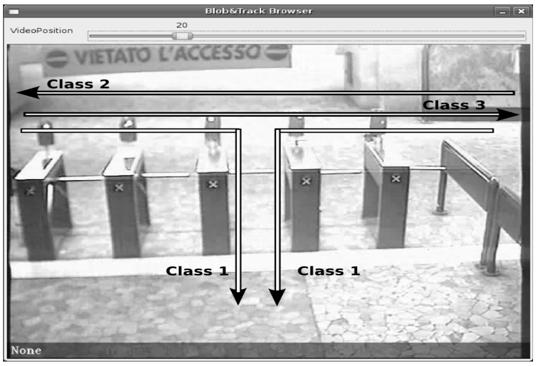

Figure 5: Definition of normal trajectories in scene.

Training set containing about 400 trajectories was created for those classes and also the testing data set with same size. The results of 1:N classification is solved by selecting the class with maximal probability (see table below).

Table 1: *Results of classification on filtered data.*

class	TP [%]	TN [%]	FP [%]	FN [%]
1	62.12	32.83	0.51	4.55
2	5.30	87.37	6.31	1.01
3	24.49	72.22	0.76	2.53

The detection of behavior defined by class 3 was explored more detailed using again ROC curve (see Figure **6** – trade-off between true positive and false positive rate).

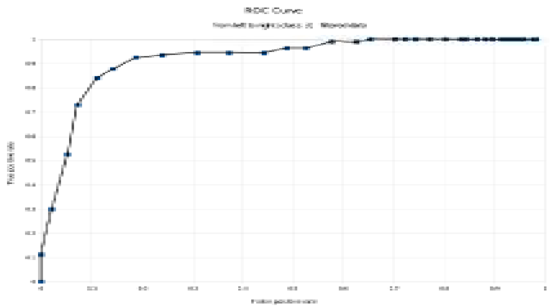

Figure 6: ROC Curve for trajectory of class 3 on filtered data.

When it is possible to define a suspect behavior by set of known trajectories, then the abnormal people behavior is recognizable from the trajectories classified as none of the know set. The behavior patterns could be also explored as feature in high level classification stage.

The performance of the trajectory pattern lookup was tested separately from tracking algorithms and the results shows, that the trajectories analysis could be deployed in the on-line system.

5. Compound event

All previous recognizers and many of other similarly based recognizers, results in one simple event. Many of such simple events itself need not necessarily mean wrong or dangerous situation. But some combination of such harmless events might results in dangerous situation. One good example of such compound event comes from station scenario. The underground or railway station operators need to prevent accidents that can happen when someone is standing too close to the edge of the platform while the train is approaching to the station.

The proposed method compounds several classes of simple events from both audio and video domain. The audio streams provide basic information about moving trains [9], but this information is not sufficient enough to make strong conclusion about possible danger. The major disadvantage is that it is hard to tell on which rail the train is coming. Therefore, also the video stream must be processed to get additional information about approaching trains. Detection of incoming trains from video-sequences is based on position analysis of segmented foreground blobs in active regions manually defined (see Figure 7 the yellow region on the rail). Other involved simple events are the active blobs (blue rectangles) intervening the area too close to the platform (red rectangles) representing the presence of the person on the platform edge.

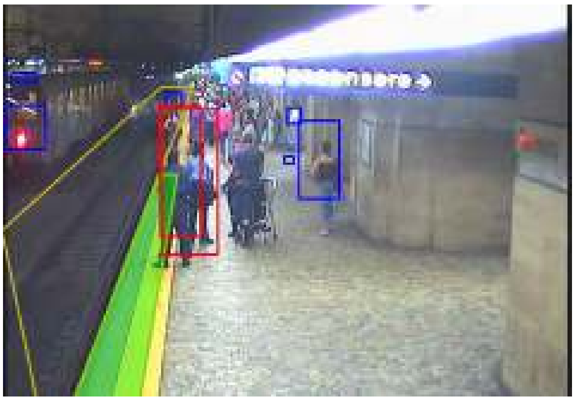

Figure 7: Simple events examples.

Having specified and detected basic events it is possible to create a complex spatio-temporal rule-based event recognizer. Ours observes three input event channels – train moving audio event, train approaching the platform event and person occurrence on the platform edge event. To successfully deal with possible weaker reliability of event detection channels the timeout intervals are defined.

The detection is formulated through the following rule (see Figure 8): if the person occurrence at the edge of the platform is detected during the period while both audio and video events are valid and both indicate moving train then the alarm is emitted.

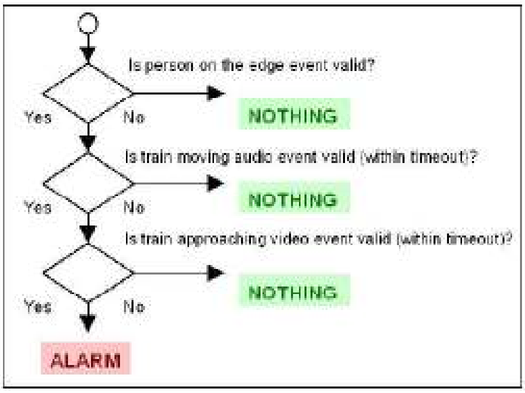
Is person on the edge event valid?
Yes
No
NOTHING
Is train moving audio event valid (within timeout)?
Yes
No
NOTHING
Is train approaching video event valid (within timeout)?
Yes
No
NOTHING
ALARM

Figure 8: Dangerous occurrence of the person on the platform edge.

Such system for compound event evaluation works on-line as far as the incorporated simple event detectors provide their results on-line. The documented example of the compound event was developed to prove the on-line concept, so we did not put much effort to create excellent simple event detectors which straightly influence the performance of the compound detector itself. Existing implementation is able to detect 76% of possible dangerous situations because it suffers of weak foreground detection. Experiments prove that substitution of this subsystem could greatly improve whole performance.

6. Evaluation

The developed methods have been applied into surveillance system working in a real site (Roma and Torino). The surveillance system was developed and implemented in the scope of the CARETAKER project [10]. The system is designed as distributed real-time agents system.

All presented recognizers were implemented into the system. The modules receive data from the sensors on-line, process them in real-time and send them back to the system. This is the most important result: the system is able to report results on-line.

The Figure 9 shows the screen-shot of the front-end application for the operator. The live video stream is augmented by processing results received from the distributed agents.

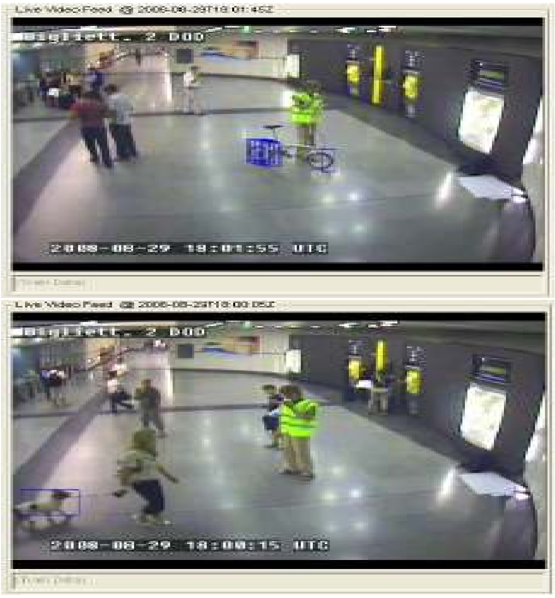
Live Video Feed @ 2008-08-29T18:01:45Z
Biglietti 2 000
2008-08-29 18:01:55 UTC
Live Video Feed @ 2008-08-29T18:00:05Z
Biglietti 2 000
2008-08-29 18:00:15 UTC

Figure 9: On-line rendering of recognition results into live video stream.

In real applications, other supporting information is taken into account. The moving object information, active regions of interest knowledge and temporal information from several consecutive frames improve the overall performance of the system.

7. Conclusions

This paper discussed several algorithms of image processing which can be used for detection of simple or compound events interesting from the point of view of CCTV surveillance in scenes similar to train stations. The main challenge of the work laid in three aspects: integration of the single-purpose detection algorithms into a complex system addressing different security threats, making the whole ensemble work in real-time and assembling the partial solutions into a proof-of-concept framework which would be functional in connection with the real CCTV surveillance system.

We have implemented modified known image and video processing methods and most of them integrated into real site system developed in CARETAKER project. The methods were evaluated on real data and we have proven that published techniques can be used in on-line surveillance systems.

The set of algorithms we have prepared demonstrates that the selected approach is feasible and leads in well functional systems.

8. Acknowledgements

This work has been supported by the Ministry of Education, Youth and Sports of the Czech Republic under the research program LC-06008 (Center for Computer Graphics), by the research project "Security-Oriented Research in Informational Technology" CEZMŠMT, MSM0021630528 and EU IST FP6 projects "AMIDA" EU-6FP-IST, IST-033812-AMIDA.

9. References

[1] C. Carincotte, X. Desurmont, B. Ravera, F. Bremond, J. Orwell, S.A. Velastin, J.M. Odobez, B. Corbucci, J. Palo, J. Cernocky. *"Toward generic intelligent knowledge extraction from video and audio: the EU-funded CARETAKER project"*, in The Institution of Engineering and Technology Conference on CRIME AND SECURITY, Imaging for Crime Detection and Prevention (ICDP), London, UK, pp.470-475, 2006.

[2] P. Zemčík, A. Herout, V. Beran, I. Potúček, O. Fučík, J. Honec, M. Richter, I. Kalová, M. Lisztwan. *Image Processing in Traffic Applications*, International Conference on Graphics, Vision and Image Processing, Cairo, 2005.

[3] P. Zemčík, A. Herout, V. Beran, J. Granát. *Hardware Accelerated Image Analysis in FPGA*, In Proceedings of Spring Conference on Computer Graphics, pp. 4, 2006.

[4] V. Beran, A. Herout, I. Řezníček. *Video-Based Bicycle Detection in Underground Scenarios*, In: Proceedings of WSCG'09, Plzeň, CZ, 2009.

[5] M. Hradiš, A. Herout, P. Zemčík. *Local Rank Patterns - Novel Features for Rapid Object Detection*, In: Proceedings of International Conference on Computer Vision and Graphics 2008, Heidelberg, DE, Springer, pp. 1-12, ISSN 0302-9743, 2008.

[6] J. Šochman, J. Matas. *Waldboost - learning for time constrained sequential detection*. In CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2, pp. 150–156, Washington, DC, USA, 2005.

[7] R. Juránek. *Detection of Dogs in Video Using Statistical Classifiers*, In: Proceedings of International Conference on Computer Vision and Graphics 2008, Heidelberg, DE, Springer, pp. 11, ISSN 0302-9743, 2008.

[8] J. Mlích, P. Chmelař. *Trajectory classification based on Hidden Markov Models*, In: Proceedings of 18th International Conference on Computer Graphics and Vision, Moscow, RU, LMSU, 2008, pp. 101-105, ISBN 595560112-0.

[9] N. Brümmer, L. Burget, J. Černocký, O. Glembek, F. Grézl, M. Karafiát, D. van Leeuwen, P. Matějka, P. Schwarz, A. Strasheim. *Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006,* IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 7, 2007.

[10] CARETAKER. http://www.ist-caretaker.org/