Event Classification for Automatic Visual-based Surveillance of Parking Lots

G.L. Foresti, C. Micheloni and L. Snidaro
Department of Mathematics and Computer Science (DIMI) University of Udine - ITALY
e-mail: foresti@dimi.uniud.it

Abstract – In this paper, a visual-based surveillance system for real-time event detection and classification in parking lots is presented. The focus is on the high-level part of the system, i.e., the event recognition (ER) module, which is able to analyze two kinds of events (i.e., simple and composite events) that occur in the observed scene. Simple events are represented by single moving objects, e.g., vehicles, pedestrians, etc. while a composite event is represented by a set of temporally consecutive simple events, e.g., people exiting a car just entered in the parking area. An adaptive high order neural tree (AHNT) is applied for recognizing both objects and complex events.

Keywords **-** Neural Trees, Object Classification, Event Recognition, Surveillance Systems

1. Introduction

Visual surveillance requires real-time interpretation of image sequences in order to automate the detection of predefined alarm situations in a given application domain [1-4]. In CCTV surveillance systems, human operators stay in a control room and look at several monitors whose images are provided by a set of multiple cameras. This solution requires the selection and storage of large amount of data and it forces the human operator to perform a repetitive task. Advanced visual-based surveillance systems provide the operator with attention focusing information which allows important events to be signaled by means of user-friendly messages and suggestions. The use of focus-of-attention messages helps the human operator to concentrate his decision capabilities on possible danger situations. In this way, possible human failures can be avoided and better surveillance performances are expected.

In the last years, some researchers have provided some solutions to the problem of event detection and scene understanding. Buxton *et al.* introduced Bayesian networks to detect interesting events into a dynamic scene and to provide interpretation of traffic situations [3]. Bobick proposed a new approach for the description and interpretation of human activities in the context of a visual surveillance task [5]. Brand *et al*. [6] proposed the use of probabilistic models, i.e., Hidden Markov Models, to capture the uncertainty of mobile object properties, while Chleq and Thonnat [7] proposed a generic framework for the real-time interpretation of real world scene with humans. Recently, Medioni et al. proposed a method for event detection in real scenes [8].

In this paper, a visual-based surveillance system for real-time event detection and classification in parking lots is presented. The focus is on the high-level part of the system, i.e., the event recognition (ER) module, which is able to analyze and classify events that occur in the observed scene. The ER module receives in input information about location, tracking and classification of mobile objects and classify an event as standard or dangerous on the basis of pre-defined object motion models. Outdoor road scenes are used as real test sites.

2. System architecture

Figure 1 shows the general architecture of the proposed visual-based surveillance system. CCD progressive color cameras are used to acquire image sequences of the monitored environment.

A change detection module (CD) [1] is applied to compare each frame *I(x,y)* of the input image sequence with a background image *BCK(x,y)*. An automatic procedure [10] is applied to find the best threshold value versus different illumination conditions. A background updating module based on the Kalman filter is used to adapt the pixels of the background image to significant changes in the scene [2]. The change detection procedure makes out a binary image *B(x,y)* where each pixel can assume two possible states: a background point or a moving object point. Groups of connected pixels, commonly called *blobs,* belonging to the class of moving points represent possible objects (e.g., vehicles, pedestrians, etc.) moving in the scene. A tracking module based on the Mean Shift algorithm [9] and using geometric information about the blobs, is applied to estimate the motion parameters of detected objects [5].

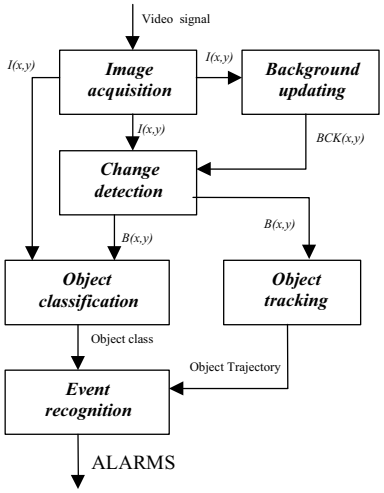


Figure 1 – General architecture of the visual-based surveillance system.

A camera calibration procedure and the ground-plane hypothesis [4] are considered to transform points from the image plane to a 2D top-view map of the scene. In particular, the map is used to display to a remote operator the position and motion parameters of each detected object. Finally, the high level part of the surveillance system is organized and applied to solve two complex problems: (a) to classify each detected blob among a predefined set of object models and (b) to understand whenever the behavior of these objects is normal or potentially dangerous. This paper is mainly focused on the high-level modules of the system: (a) object classification and (b) event recognition.

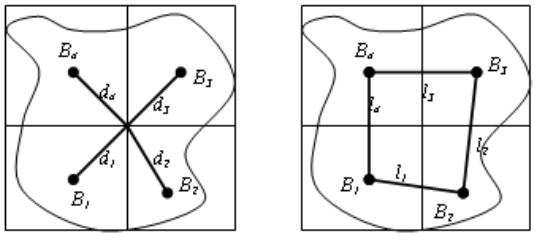
3. Object Classification

The object classification module is composed of a new classifier, called adaptive high order neural tree (AHNT) [11]. The AHNT is a hierarchical multi-level neural network, in which the nodes (first-order or a high order perceptron) are organized into a tree topology. It successively partitions the training set into subsets, assigning each subset to a different child node. First order perceptrons split the training set by hyperplanes, while n-order perceptrons use n-dimensional surfaces. The AHNT is learned by feature vectors extracted from 2D detected blobs of interesting objects on the image plane. In particular, the blob region B is divided into four parts, *q1, q2, q3, q4*, called quartiles [11], according to the position of its center of mass (Figure 2a).

The four distances, d_1, d_2, d_3, d_4, between the center of mass of the blob B and the center of mass of the four quartiles (see Fig. 2a) are computed. In order to increase the robustness of such a representation, the distances l_1, l_2, l_3, l_4, between the center of mass of the four quartiles have been considered (Fig. 2b). The pattern $\boldsymbol{p}(B)$, related to the blob B, is composed by the following eight values:

$$\boldsymbol{p}(B)=[d_1,d_2,d_3,d_4,l_1,l_2,l_3,l_4] \quad (1)$$

The object classification obtained on N consecutive frames of the input sequence is analyzed by a Winner-takes-all classification procedure [2] which computes the final classification of the detected object.



(a) *(b)*

Figure 2 – The four quartils and the related distances from center of mass of the blob. (b) The four distances l_i.

The class which receives the maximum number of occurrences is selected as the class of the detected object. The matching process necessary to associate the same object over consecutive frames is performed by the tracking module on the basis of a cost function taking into account the object motion constraints, e.g., maximum speed, direction, etc.

4. Object Tracking

The tracking module applies the Mean Shift algorithm [9] integrated with geometric information about the size of the blob on the image plane and its position on the 2D top view map. In particular, the ratio $r=h/w$ between the height and the width of the minimum bounding rectangle of the blob and its position $\boldsymbol{pos}(x,y)$ on the 2D top-view map have been considered. When one or more objects have been detected by the CD module, the Mean Shift algorithm is applied to find the best blob correspondences between the current and the previous frame.

Let $O_i(t)$ $i=1,..,n$, be the set of n objects detected at the time instant t, and let $O_{ij}(t-1)$ $j=1,..,m$, be the set of m objects detected at the time instant $(t-1)$ that satisfy the above constraints with the object $O_i(t)$. Then, let $[\boldsymbol{pos}_i(x,t), \boldsymbol{pos}_i(y,t)]$ be the coordinates in pixels of the real position of the object $O_i(t)$ at the time instant t, and let $[pos_{ij}(x,t), pos_{ij}(y,t)]$ be the coordinates in pixels of the real position of the candidate objects. The matching function $MF_{ij}(t)$ is represented by the Euclidean distance between the coordinate points of the candidate objects.

Since, the tracking module extracts from each image an estimated measure $\boldsymbol{pos}_i(x_t,y_t)$ of the position of the object on the ground plane, the trajectory analysis module

processes a set of N consecutive object positions on the ground plane, $\boldsymbol{pos}_i(x_{t-N},y_{t-N})$, $\boldsymbol{pos}_i(x_{t-N+1},y_{t-N+1})$, ..., $\boldsymbol{pos}_i(x_t,y_t)$. At each time instant t, the module computes the displacement of the tracked object on the 2D map.

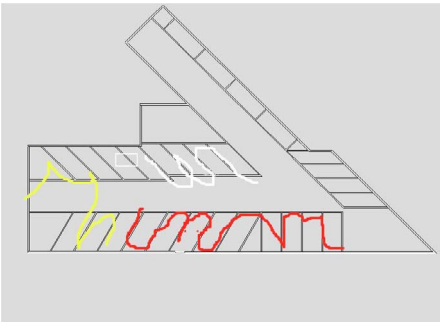


Figure 3 – Some examples of atypical trajectories of pedestrians in a parking lot.

The approximation of the trajectory of a moving object is performed by employing the parametric Bezier curve as follows:

$$\mathbf{P}(p)=\sum_{i=0}^{N}\boldsymbol{pos}_i\left(x_{t-N+i},x_{t-N+i}\right)\mathbf{J}_{n,i}(p)\quad 0\le p\le 1 \qquad (2)$$

where $\mathbf{J}_{n,i}(p)=\binom{n}{i}p^i(1-p)^{n-i}$. Figure 3 shows some atypical trajectories of pedestrians on the 2D top view map of a parking area. Atypical trajectories refer to pedestrians that move several times around to one or more vehicles.

5. Event Recognition

An event is characterized by a set of k classified objects over a sequence of n consecutive frames, $\boldsymbol{E}(k,n)$, $k=1,..,K$ and $n=1,..,N$. The goal of the proposed system is to detect an event and classify it into a set of tree levels of increasing dangerousness: (a) normal event, (b) suspicious event and (c) dangerous event. Suspicious and dangerous event generates two kinds of different alarm signals that are displayed on the man-machine interface of the remote operator. Two types of events have been considered: (a) simple events and (b) composite events.

In the considered environment, i.e., a parking lot, a simple event is represented by a vehicle moving and/or stopping in allowed areas or a pedestrian walking with typical trajectories (e.g., almost rectilinear for a long number of frames). Suspicious events are represented by pedestrians walking with trajectories not always rectilinear, pedestrians stopping in a given parking area, pedestrian moving around a vehicle, etc. Dangerous events are represented by pedestrians or vehicles moving or stopping in not allowed areas, pedestrians moving with atypical trajectories (e.g., moving around several vehicles as shown in Figure 3).

An off-line Event Database (ED) is built with the models of normal, suspicious and dangerous simple events. For each event class, a set of features is extracted and stored in the ED. In particular, the class C_i of the detected object and the N coefficients of the parametric Bezier curve which approximates the object trajectory are used as pattern to learn an AHNT.

An Active Event Database (AED) is built containing info about the detected simple events. For each event, the initial and final instants, t_{in} and t_{fin}, the class C_i of the detected object, the N coefficients of the parametric Bezier curve which approximates the object trajectory, the average object speed S_i, the initial and final object position on the 2D top view map $\boldsymbol{pos}_i(x_{fin},y_{fin})$ and $\boldsymbol{pos}_i(x_{tfin},y_{tfin})$ are stored. An age counter is used to eliminate from the AED old events. The AED is inspected at a given time instant to analyze the active events and verify if there are events having some correlations. For example, if the AED contains some simple events such as people moving in the parking area (i.e., a people starting from a given point of the parking and stopping in a different point or exit from the parking) and a vehicle moving in the parking area (i.e., a vehicle entering/leaving the parking), an automatic procedure checks if some of these events are temporally and spatially correlated. If there exist two or more temporally consecutive simple events with initial or final position spatially closed, a composite event is generated and stored in the AED. An example of a composite event in the context of a parking lot can be represented by a sequence of simple events as the following: (a) a vehicle entering the parking area, (b) moving with a given trajectory, (c) stopping in a given position, (d) a person exits from that vehicle, (e) moving in the parking area and (f) exits the parking area. When a composite event has been detected, its classification is easily performed by analyzing if it is contained into a set of models of normal or unusual composite events previously defined by the system operator.

6. Experimental results

The proposed system has been tested on real image sequences taken from a parking lot. The length of the considered sequences ranges from 20 seconds (400 images at a rate of 20 images per second) to 5 minutes. Various scenarios have been considered and, for each of them, ground truth data (normal, suspicious or dangerous events) have been manually defined. Figure 4 shows some results of the proposed system

An AHNT trained with about 50 patterns representing normal events, 30 patterns representing suspicious events and 20 patterns representing dangerous events for both vehicle and pedestrian object models has been built. The

obtained AHNT is composed of 41 nodes distributed over 4 levels. In particular, 26 nodes are standard (first-level) perceptrons, 11 are second-order perceptrons and the remaining are third-order perceptrons. Performances of the proposed system have been measured in terms of correct object classification and in terms of false alarms and misdetection errors in the event recognition process. Table A shows the percentage of correct object classification and the distribution in other classes of the bad classification obtained on a large test set (about 10^3 images).

	CAR	***CYCLES***	***PEDESTRIAN***
CAR	98	-	-
CYCLES	1	9	82
PEDESTRIAN	3	78	11

Under normal conditions the mean occurrences of false alarms are equal to 7 while the missed alarms to 3. In context of bad conditions these two parameters are respectively equal to 18 and 13.

Acknowledgements

This work was partially supported by the Italian Ministry of University and Scientific Research (MIUR) within the project "Distributed systems for multisensor recognition with augmented perception for ambient security and customization" .

References

[1] G.L. Foresti, C.S. Regazzoni and R. Visvanathan, *Special Issue on Video Communications, Processing and Understanding for Third Generation Surveillance Systems*, *Proceedings of IEEE*, Vol. 89, no. 10, October 2001.

[2] G.L. Foresti, C.S. Regazzoni and P. Varnshey, *Multisensor Surveillance Systems: The Fusion Perspective*, Kluwer Academic Publishers, Norwell, MA, USA, 2003.

[3] H. Buxton and S. Gong, "Visual surveillance in a dynamic and uncertain world", *Artificial Intelligence*, Vol. 78, No. 1-2, 1995, pp. 431-459.

[4] R.T. Collins, A.J. Lipton, H. Fujiyoshi, T. Kanade, "A system for video surveillance and monitoring", *Proceedings of IEEE*, Vol. 89, no. 10, October 2001, pp. 1456-1477.

[5] A.F. Bobick, "Computer seeing action", in *Proc. of the 7th Annual British Machine Vision Conf.e,* 1996, pp. 13-22.

[6] M. Brand, N. Oliver and A. Pentland, "Coupled hidden Markov models for compelx action recognition" in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, USA, June 1997.

[7] N. Chleq and M. Thonnat, "Real-time image sequence interpretation for video-surveillance applications" in *Proc. of IEEE Int. Conf. on Image Processing,* Lausanne (CH), September 1996, pp. 801-804.

[8] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia, "Event Detection and Analysis from Video Streams", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, 2001, pp. 873-889.

[9] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, 2002, pp. 603 - 619.

[10] P.L. Rosin, "Thresholding for Change Detection", *Computer Vision and Image Understanding*, Vol. 86, no. 2, pp. 79-95, 2002.

[11] G.L. Foresti and T. Dolso, "Adaptive High-Order Neural Trees for Pattern Recognition", *IEEE Trans. on System, Man and Cybernetics-Part B*, vol. 34, no. 2, 2004, pp. 988-996.

[12] K.J. Bradshaw, I.D. Reid, and D.W. Murray, "The active recovery of 3D motion trajectories and their use in prediction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, 1997, 219-234.

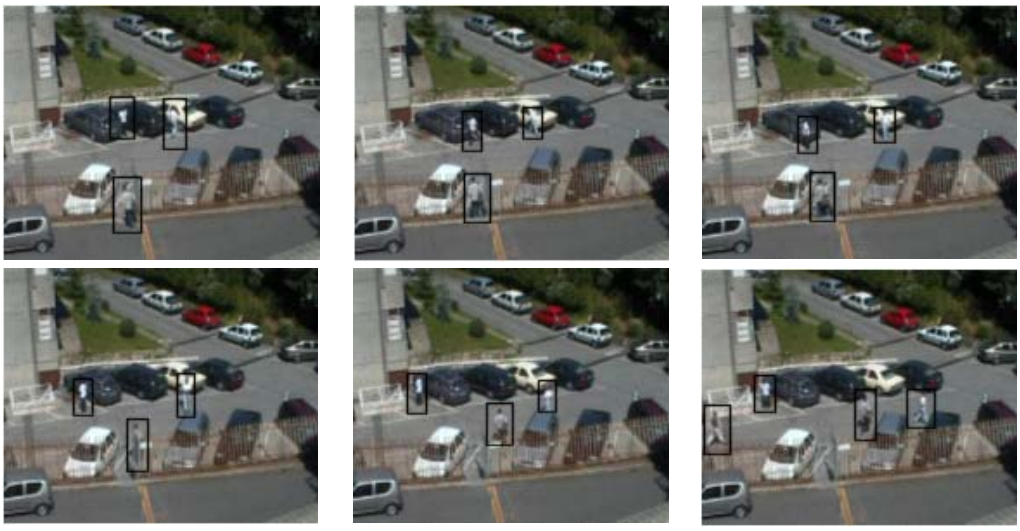


Figure 4 – Some examples of suspicious events in a parking lot