VEHICLE EXTRACTION BASED ON FOCUS OF ATTENTION, MULTI FEATURE SEGMENTATION AND TRACKING

Andrea Cavallaro[*], *Francesco Ziliani*[*], *Roberto Castagno*[**], *and Touradj Ebrahimi*[*]
[*] Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland
[**] Nokia Mobile Phones, Tampere, Finland
Tel: +41 21 693 2708; fax: +41 21 693 7600
e-mail: andrea.cavallaro@epfl.ch

ABSTRACT

In this paper, we propose an automatic object tracking method for traffic video surveillance. The method is designed as a hybrid between a region based and a feature based technique. In a first stage, a motion detection algorithm identifies the objects from the background providing binary masks of the moving objects. In a second stage, a segmentation tool based on a multi-feature analysis further segments the areas corresponding to moving objects into homogenous regions. For each region, the method provides a set of characteristic feature values, which are used to track the regions (and thus the objects) along time. The results of this low-level analysis can be exploited by the content understanding module of an advanced video surveillance system for the detection of potentially dangerous situations, for law enforcement purposes, and for statistical traffic analysis.

1 INTRODUCTION

The evolution of advanced video surveillance systems has lead from the so called first generation CCTV systems to the second generation PC based systems [3]. This favored the introduction of automatic digital image processing techniques to assist an human operator in video surveillance tasks, thus reducing the need for his continuous attention. The user intervention can therefore be limited to higher level tasks. The user interprets critical situations, makes decisions in doubt cases, and chooses the most appropriate action when an alarm is generated.

An automatic tool for video surveillance should achieve the extraction of the information of interest from the video input and it should provide reliable data to a content understanding module. The basic structure of such a system is depicted in Fig. 1. Here, the initial visual data is precessed so as to pass to the content understanding step a reduced amount of data which provides the most significant information on the scene content.

The detection of the areas where a change in the scene has occurred with respect to a reference frame is already an appropriate solution for some traffic surveillance applications. This is true, for example, when the goal of the application is an approximate vehicle counting or an estimation of traffic flow conditions. In this case the video-based surveillance system is used as an alternative to physical detectors, such as magnetic loops. This simple solution is uneffective if a more complex analysis of the traffic conditions is required.

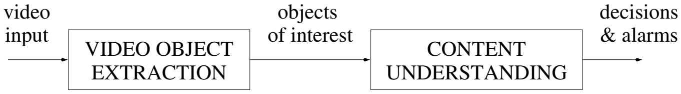

Figure 1: Block diagram of an advanced video surveillance system. The video input is processed in order to provide to a content understanding module only the objects of interest, which are moving in the scene

We propose an approach to video surveillance based on the concept of *focus of attention*. The attention of the system is driven by a change detection mask, that is defined by a statistical approach. This provides a binary mask, which allows us to delimit the areas of interest in the scene. These areas are then segmented into spatio-temporal homogeneous regions, which are tracked along the sequence. This procedure provides a trajectory for each region that may help in automatically recognizing alarm situations.

The paper is organized as follows. Section 2 reviews the different tracking methods that are currently used in the framework of traffic surveillance. The proposed approach is presented in Sec. 3. In Sec. 4 we present the experimental results and, finally, in Sec. 5 we draw the conclusions.

2 STATE OF THE ART

An advanced video based surveillance system is required to monitor objects in a scene. The monitoring implies that the different objects are tracked along time. In the following we present a brief overview of tracking methods, which have been proposed in the framework of traffic surveillance.

2.1 Region based tracking

A region-based method tracks blobs of pixels, which roughly correspond to vehicles. To achieve this, it relies on information (such as motion, color, and texture properties) provided by the entire region [4]. This approach works properly only in case of free flow traffic and usually requires that the camera is placed in a high position with respect to the road. In case of congested traffic conditions or cameras placed in a low position with respect to the road plane, in fact, vehicles partially occlude each other, thus leading the method to group together more vehicles in a unique large blob. This causes the loss of the track of each single vehicle.

2.2 Active contour based tracking

Active contour models (snakes) rely on the information provided by the object boundaries [6, 7]. A contour-based representation can help in reducing the computational complexity. Furthermore, it allows the tracking of both rigid and non-rigid objects. On the other hand it is unable to track vehicles that are partially occluded. To overcome the problem of partial occlusions, in [8] a Kalman filtering approach and optical flow measurements have been introduced in the active contour model.

2.3 3D model based tracking

The problem of tracking vehicles which are partially occluded can be solved by considering their 3D models [5]. The definition of parametrized vehicle models make it possible to exploit the a priori knowledge about the shape of typical objects in traffic scene. This approach is computationally intensive and it presents two major drawbacks: the need of object models with detailed geometry for all vehicles that could be found in the scene, and the lack of generality. This second drawback does not allow the system to detect objects different from vehicles. For example, when the monitored scene is a highway, the detection of people and animals is important for interpreting the scene in case of dangerous situations, but with such a method they cannot be detected.

2.4 Feature based tracking

Using a feature based method, instead of the entire object, its sub-parts are tracked [1]. An example of these sub-parts is represented by the corners of the objects, which allow tracking also in case of partial occlusions. A major drawback of such an approach is the problem of grouping the features by finding which of them belong to the same object.

By considering these previous works, the approach we propose has been designed as an hybrid between the region based and the feature based techniques. It exploits the good characteristics of the two by considering first the object as an entity and by tracking then its sub-parts.

In a first stage a motion detection algorithm identifies the objects from the background and provides a mask defining the areas of the image containing the moving objects. In the second stage, objects are tracked by projecting the representatives (centroids) of the regions obtained with a multi-feature segmentation. The details of the method are presented in the following section.

3 PROPOSED METHOD

The proposed method for advanced surveillance segments automatically the veichles from the background, and it has the ability of dealing with different traffic conditions and variety of objects.

3.1 Segmentation driven by focus of attention

The proposed approach to video objects extraction and tracking for traffic surveillance is based on the concept of *focus of attention*. The *attention* of the system is driven by a change detection mask, defined by the statistical approach described in [9]. This provides a binary mask, which allows us to delimit the areas of interest in the scene as shown in Fig. 2. The areas of interest are represented as a matrix with the same size of the image. Every element of the matrix can take one of the two possible values depending on whether the pixel has been detected as changed or not by the motion detector.

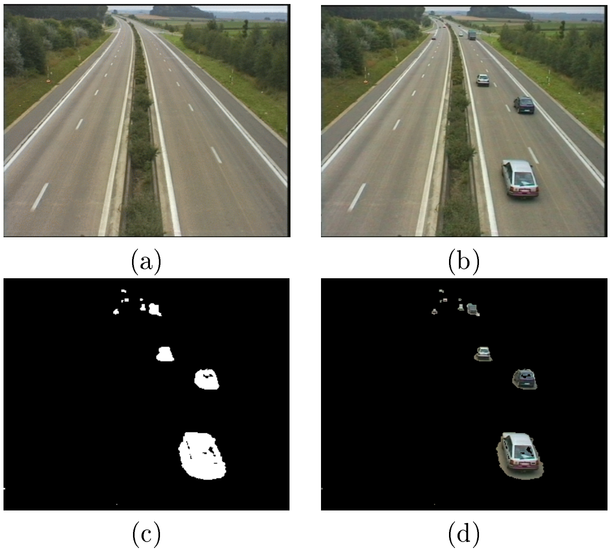

Figure 2: Traffic surveillance sequence highway n25w22 (courtesy of European ACTS project 304 Modest): (a) reference frame; (b) frame n.215; (c) corresponding change detection mask; (e) moving objects defined by the change detection mask (here the change detection mask is superposed on the original frame).

Only the areas of interest are considered by the following step, which takes into account the spatio-temporal properties of the pixels in the changed areas and extracts spatio-temporal homogeneous sub-regions. These sub-parts of the objects are clusters obtained with the Fuzzy C-Mean algorithm as proposed in [2]. The clustering is performed on a multi-feature space composed by speed, position, color and texture of the changed pixels. Each object is processed separately and it is decomposed in a set of non-overlapping clusters or regions as shown in Fig. 3.

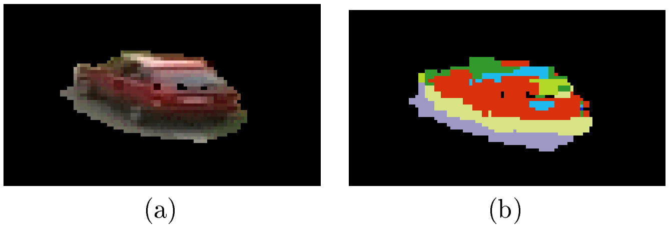

Figure 3: Example of segmentation driven by focus of attention: (a) area of interest defined by the change detection mask and (b) regions defined by the multi feature segmentation.

The properties of each region are represented in a centroid summarizing the values of speed, position, color and texture of the cluster. To link these properties to the correct object in each frame, a multi tracking procedure is performed.

3.2 Automatic object tracking

The proposed tracking method integrates the information along the sequence and provides a trajectory for each region: this may help in automatically recognize alarm situations. One of the advantages of this approach is the ability to handle partial occlusions, and appearing and disappearing of objects from the scene.

Given the objects of interest in the new frame, the objects of interest in the current frame and their segmentation in spatio-temporal regions, the proposed tracking procedure performs two different tasks. First, it defines a correspondence between the objects of interest in the current frame n and those detected in the new frame $n+1$. Second, it provides an effective initialization for the segmentation procedure of each object in the next frame $n+1$. This initialization implicitly defines a correspondence between the regions in frame n and the regions in frame $n+1$.

In order to explain in detail our technique, we limit the discussion to a single object of interest detected in the current frame. The same procedure is repeated for all the objects in the frame. Let us refer to an object of interest as $O_k(n)$. As explained above, $O_k(n)$ is segmented with the Fuzzy C-Mean algorithm in r spatio-temporal regions $R_{1,k}(n), ..., R_{r,k}(n)$. The values given to r depend on the size of the object $O_k(n)$ and they have been chosen empirically. Each region is projected through motion compensation on frame $n+1$. This operation, referred to as *centroid projection*, is performed by adding to the position values of a centroid its estimated vertical and horizontal displacement. We then compute the mean of the gravity centers of all the projected regions belonging to the same object $O_k(n)$. This coordinate, c_k, is compared with the gravity center of all the objects detected with the statistical change detection in the frame $n+1$. We define a correspondence between the object $O_k(n)$ and the object $O_j(n+1)$ whose gravity center has the minimum Euclidean distance with c_k. Once the correspondence has been performed the object $O_j(n+1)$ is segmented by applying the Fuzzy C-Mean algorithm. The initialization of this segmentation process is defined by the projected centroids of the r regions $R_{1,k}(n), ..., R_{r,k}(n)$. Thus a natural correspondence between the regions of the object $O_k(n)$ and the regions of the object $O_j(n+1)$ is obtained.

The overall video object extraction and tracking scheme is summarized in Fig. 4.

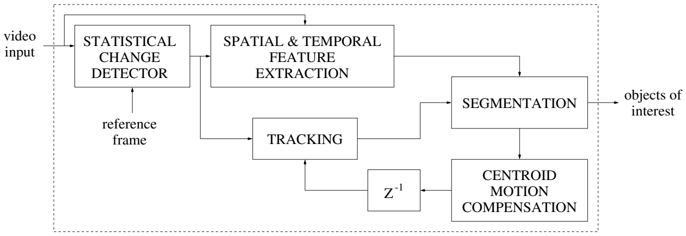
video input
STATISTICAL CHANGE DETECTOR
reference frame
SPATIAL & TEMPORAL FEATURE EXTRACTION
TRACKING
SEGMENTATION
objects of interest
Z^-1
CENTROID MOTION COMPENSATION

Figure 4: Block diagram of the proposed video object extraction and tracking mechanism

4 RESULTS

The proposed method has been tested on typical outdoor surveillance sequences. Four representative frames of MPEG-7 surveillance test sequence have been displayed in the first column of Fig. 5. The sequence is shot in daylight conditions from a camera placed on a bridge passing above a highway. The second column shows an example of the automatic object extraction and tracking capabilities of the proposed algorithm. The vehicle is automatically extracted and tracked along the frames after the statistical change detector has provided the mask containing all the moving objects. The reported example shows the tracking of a single vehicle from its appearing in the camera scope in frame 100, until it exits the scene in frame 130. In the same way, all the other objects in the scene are automatically extracted and separately tracked along the frames, thus providing the content understanding module with segmented objects and their associated trajectories. This information will help this last module in describing events in the scene and in generating alarms in case of dangerous situations.

The extraction of the vehicle is not precise due to the presence of shadows that are detected as moving objects. To overcome this problem an approach that

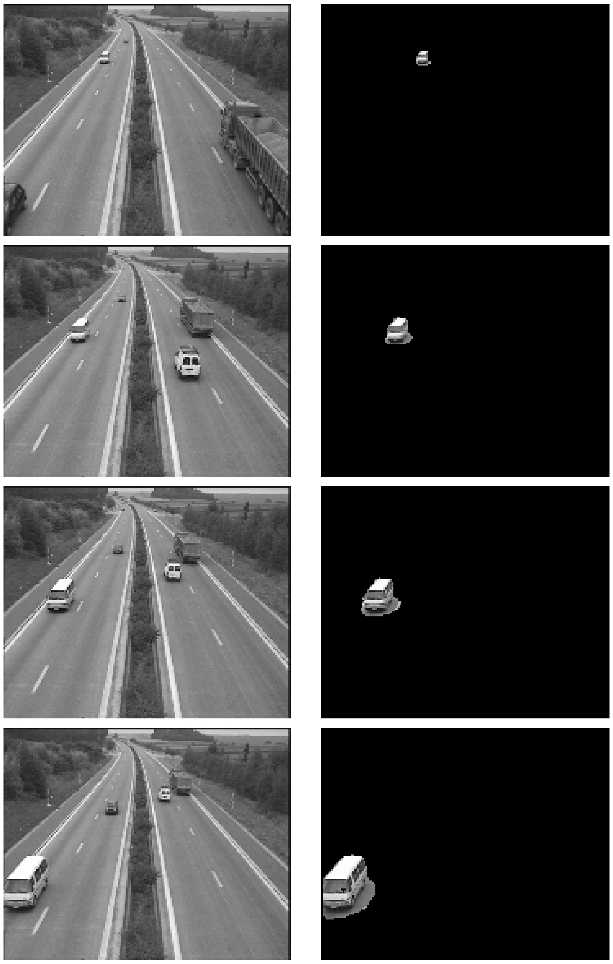

Figure 5: Traffic surveillance sequence highway n25w22 (courtesy of European ACTS project AC304). *First column*: original frames n.100, n.110, n.120, n.130; *second column*: example of object tracking. The vehicle has been automatically extracted and tracked

consider the local color properties of the regions can be used as proposed in [9].

5 CONCLUSIONS

We presented an efficient method for extracting vehicles in traffic surveillance sequences. The proposed approach to object tracking is an hybrid between a region based and a feature based technique. The efficiency is achieved by considering only the regions in the images that are interesting for the application (moving objects). The tracking is based on the sub-parts of the objects, identified by a multi-feature segmentation, leading to the flexibility of the technique.

The procedure is fully automatic and provides throughout the sequence a trajectory for each object. This helps the following content understanding module to describe events in the scene, and to automatically recognize alarm situations.

References

[1] D. Beymer, P. McLauchlan, B. Coifman, J. Malik, "Real-time Computer Vision System for Measuring Traffic Parameters," Proceedings of Computer Vision and Pattern Recognition, San Juan, Puerto Rico, June 1997, pp. 495–501.

[2] R. Castagno, T. Ebrahimi, M. Kunt, "Video Segmentation based on Multiple Features for Interactive Multimedia Applications," IEEE Trans. on Circuits and System for Video Technology, Vol. 8, No. 5, pp. 562–571, September 1998.

[3] C. Sacchi and C.S. Regazzoni, "Multimedia Communication Tecnhiques for Remote Cable-Based Video-Surveillance Systems," Proceedings of the 10th International Conference on Image Analysis and Processing (ICIAP), Venice, Italy, September 1999, pp. 1100–1103.

[4] K.P. Karmann and A. von Brandt, "Moving object recognition using an adaptive background memory," Time-Varying Image Processing and Moving Object Recognition (V. Capellini Ed.), Elsevier, The Netherlands, 1990.

[5] D. Koller, K. Danilidis and H.H. Nagel, "Model-Based Object Tracking in Monocular Image Sequences of Road Traffic Scenes," Int. Journal of Computer Vision, 10:3, pp. 257–281, 1993.

[6] D. Koller, J. Weber and J. Malik, "Robust Multiple Car Tracking with Occlusion Reasoning," Proceedings of European Conference on Computer Vision (ECCV), pp. 189–196, 1994.

[7] N. Paragios and R. Deriche, "Geodesic Active Regions for Motion Estimation and Tracking," Proceedings of the 7th International Conference on Computer Vision (ICCV), 1999.

[8] N. Peterfreund, "Robust tracking of position and velocity with kalman snakes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, June 1999.

[9] F. Ziliani and A. Cavallaro, "Image analysis for video surveillance based on spatial regularization of a statistical model-based change detection," Proceedings of 10th Int. Conference on Image Analysis and Processing (ICIAP), Venice, Italy, September 1999, pp. 1108–1111.