AUTOMATIC TRACKING OF MULTIPLE PEDESTRIANS WITH GROUP FORMATION AND OCCLUSIONS

PEDRO MENDES JORGE
Departamento de Engenharia da Electrónica e Telecomunicações e de Computadores
Instituto Superior de Engenharia de Lisboa
Rua Conselheiro Emídio Navarro, nº1
1940-014 Lisboa, Portugal
email: pmj@isel.ipl.pt

ARNALDO J. ABRANTES
Departamento de Engenharia da Electrónica e Telecomunicações e de Computadores
Instituto Superior de Engenharia de Lisboa
Rua Conselheiro Emídio Navarro, nº1
1940-014 Lisboa, Portugal
email: aja@isel.ipl.pt

JORGE S. MARQUES
Instituto Superior Técnico
Instituto de Sistemas e Robótica
Av. Rovisco Pais
1049-001 Lisboa, Portugal
email: jsm@isr.ist.utl.pt

ABSTRACT
This work addresses the problem of automatic tracking of pedestrians observed by a fixed camera in outdoor scenes. Tracking isolated pedestrians is not a difficult task. The challenge arises when the tracking system has to deal with temporary occlusions and groups of pedestrians. In both cases it is not possible to track each pedestrian during the whole video sequence. However, the system should be able to recognize each pedestrian as soon as he/she becomes visible and isolated from the group. This paper presents methods to tackle these difficulties. The proposed system is based on a hierarchical approach which allows the application of the same methods for tracking isolated pedestrians and groups.

KEY WORDS
Image Processing, Video Surveillance, Pedestrian Tracking

1. Introduction

Tracking of pedestrians is a stimulating problem from a theoretical point of view and it has important applications in surveillance, human-machine interface and gesture analysis. Many works have attempt to accurately estimate the shape of the human body along a video sequence using 2D and 3D models of the object boundary (e.g., see [1, 2]). In surveillance applications the objects have a small area in the image and such a detailed description is not available. Region based methods are used instead. These methods usually rely on the detection of active regions on the image, followed by tracking algorithms used to match the detected regions in consecutive frames [3, 4, 5]. Sophisticated methods have been proposed for detection of active regions e.g., based on background modelling [6, 7]. The main difficulty concerns the correspondence problem in the presence of multiple interacting objects. Particulary, two difficulties are the occlusion of pedestrians by the scene and by other pedestrians. In both cases it is not possible to track the pedestrians while they are occluded. However in some applications, the long term tracking of each pedestrian is an important goal i.e., the tracking system should be able to recognize and track each pedestrian as soon as he/she becomes visible after being occluded. This paper describes methods to tackle these difficulties.

2. Tracking System

Given an image sequence we wish to track all the pedestrians appearing in the scene during a given time interval. It is usually simple to track pedestrians in the case that they are not occluded. When occlusions occurs it is not possible to track the pedestrian. However we would like to recognize each pedestrian after he/she becomes visible and isolated. A system based on three steps (Fig. 1) will be used in this paper.

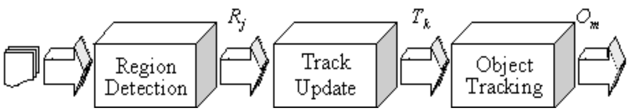


Figure 1. Block diagram of the tracking system.

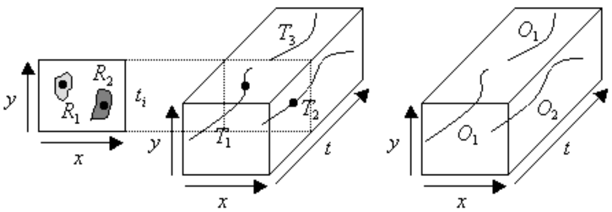


Figure 2. Example of the output of the three steps.

The first step computes the active region in each frame. An active region is a connected subset of pixels in which motion is observed. The second step computes the evolution of each region along time when no occlusion is observed. The trajectories of the region centroids are denoted as *tracks*. Figure 2 shows a set of 3 tracks detected in a video sequence. Finally, the third step tries to match all the tracks associated to the same pedestrian or group of pedestrians. We note that all the three steps apply in the case of isolated pedestrians as well as in the case of pedestrian groups. When a pedestrian joins or leaves a group, new tracks are generated.

3. Data Representation

Three concepts are used to represent the image data, using a bottom-up approach: *active region* (connected subset of active pixels); *track* (sequence of matched regions detected in consecutive frames) and *object* (sequence of tracks associated to the same pedestrian or group). Considering that pedestrians can occur isolated or in group, the object can be *simple* or *compound*, respectively. An example is shown in figures 3 and 4.

A set of images from a video sequence (figure 3) and a x-t (figure 4) plot are presented. In the x-t plot each region at frame t is represented by the x coordinate of its center. The line segments represent the tracks. The objects O_1 and O_2 are associated with

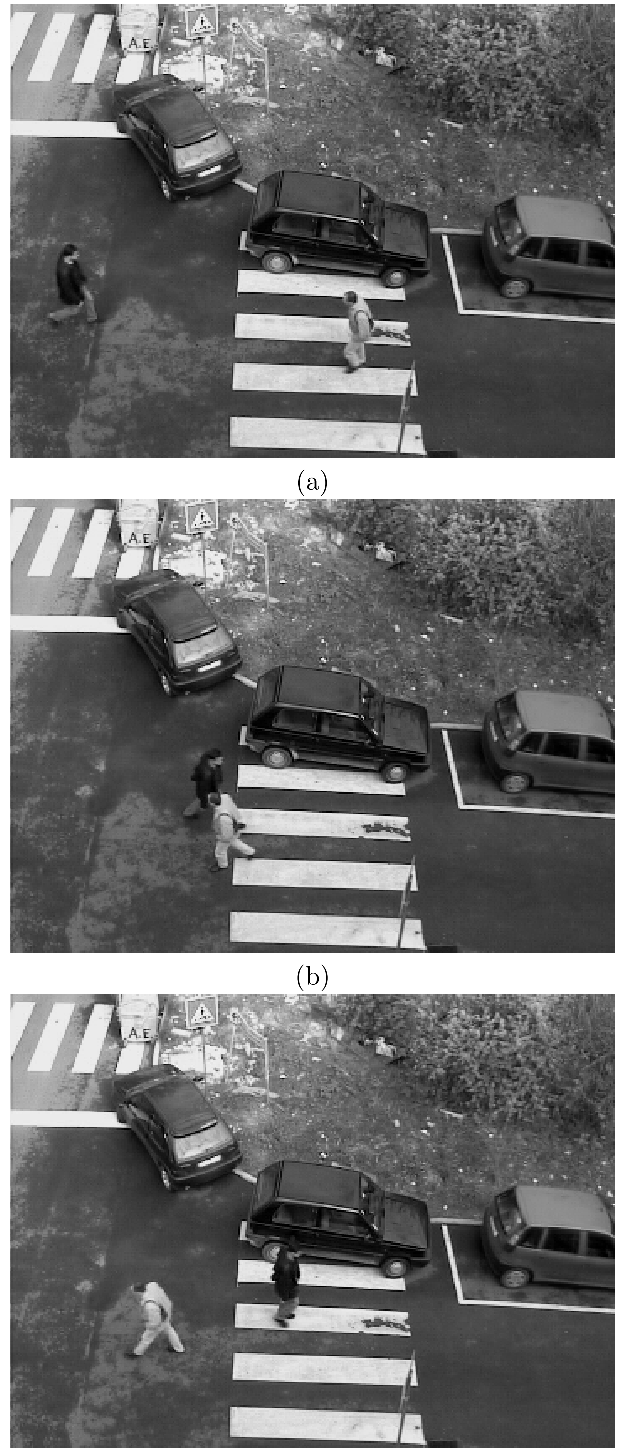


Figure 3. Occlusion example. Frames before, during and after the occlusion (frame number 60 (a), 99 (b) and 120 (c)).

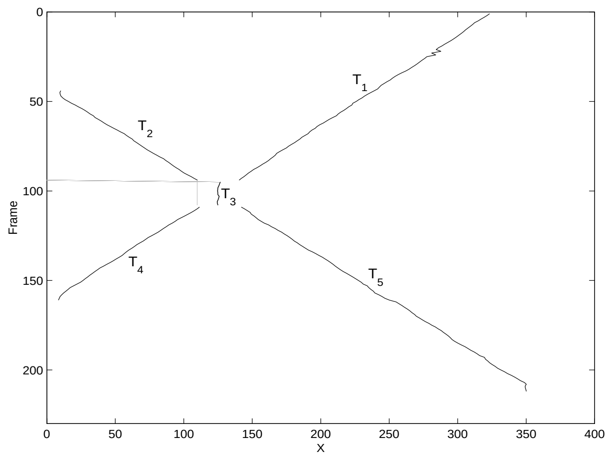


Figure 4. x-t plot of the occlusion example of figure 3.

tracks $\{T_1, T_4\}$ and $\{T_2, T_5\}$, respectively, and represent isolated pedestrians (simple objects). The object O_3 is associated with track T_3 and represents a group of pedestrians (compound object) that occurs when the pedestrians overlap in the image.

Let a_i denote an active pixel. It is assumed that a_i is defined as follows

$$a_i = \{[x_i y_i], [r_i g_i b_i]\} \tag{1}$$

where $[x_i y_i]$ is a vector of coordinates (pixel column and line) and $[r_i g_i b_i]$ is a vector of color components. An active region R_j^t detected at t-th frame is a set of active pixels

$$R_j^t = \{[a_{j1} ... a_{jN_j}], \theta_j\} \tag{2}$$

where a_{jk} is an active pixel belonging to R_j^t and θ_j is a vector of parameters characterizing the region properties. For example, θ_j may contain shape and color information.

A track is defined by

$$T_k = \{[R_i^{t_{k1}} R_j^{t_{k1}+1} ... R_l^{t_{k2}}], \zeta_k\} \tag{3}$$

where $[R_i^{t_{k1}} R_j^{t_{k1}+1} ... R_l^{t_{k2}}]$ is a sequence of matched regions detected between frames t_{k1} and t_{k2} and ζ_k defines the track parameters (e.g., color distribution, average velocity).

An object represents either an isolated pedestrian or a group of pedestrians. In the first case, it is denoted as a *simple object* and in the latter case, it is denoted as a *compound object*. A simple object is a sequence of tracks which corresponds to the position of the same pedestrian along the video sequence. Therefore, it is represented by

$$O_m = \{[T_q ... T_u]\} \tag{4}$$

A compound object is also a sequence of tracks but it also includes the indices of the objects which belong to the group. In that case,

$$O_m = \{[T_q ... T_u], I_m\} \tag{5}$$

where I_m contains the indices of the objects (simple or compound) which joined the group. Every time a person joins or leaves the group a new compound object is created.

4. Tracking Methods

This section describes the methods used in each block of the tracking system: region detection, track update and object tracking. To detect the active pixels, a background subtraction approach is used. The intensity of each pixel is compared with a background image. The difference between each pixel and the corresponding pixel of the background image is computed. The pixel is classified as active when the difference exceeds a threshold. Then, morphological close is applied to fill small gaps. The connected sets of active pixels are considered as active regions.

Track update (block 2) is performed by using a mutual choice criterion. A region R_j^t is associated to a track T_k^{t-1} if R_j^t is the closest region and if T_k^{t-1} is the track closest to the region. Assuming that $T_k^{t-1} = \{[R_i^{t_{k1}} ... R_l^{t-1}], \zeta_k^{t-1}\}$, the updated track is $T_k^t = \{[R_i^{t_{k1}} ... R_l^{t-1} R_j^t], \zeta_k^t\}$ where ζ_k^t is an updated version of the track parameters. After this procedure, all the tracks which are not updated end and all the unmatched regions are considered as the beginning of new tracks.

Object update is performed when new tracks are created. New tracks are detected in three different cases: i) a new pedestrian appears in the scene; ii) a pedestrian which was occluded by the scene becomes visible and iii) pedestrians join or leave a group. In the first case, a new simple object is created. In the second case, the new track is associated to an existing object which represents the same pedestrian along the video sequence (see O_1 in figure 2). The last case is the most complex. Every time a person joins a group a new object (compound) is created (see figure 5). The parameter I is initialized with the labels of all the objects belonging to this new compound object. When a person leaves the group, its track is associated to an existing object and a new compound object with less members (new group) is usually created.

5. Experimental Results

The tracking system was applied to video sequences which illustrate all the situations described above. The sequences were acquired with a digital color camera Canon-MV30i at 25 frames/seg. Figures 6-8 show

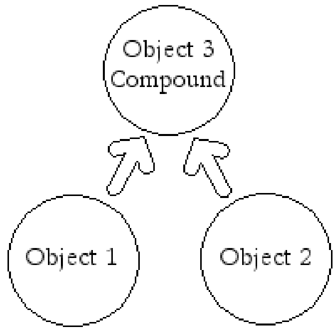


Figure 5. Compound object creation with the information of the belonging objects.

three examples. In each case we show one frame of the video sequence and the tracking results obtained with the proposed algorithm described in this paper. Figure 6 illustrates the occlusion of a pedestrian by the scene. The algorithm manages to associate all the detected tracks to a single object. Figure 7 illustrates the occlusion of pedestrians walking in opposite directions. A compound object $O_3 = \{T_3, I_3\}, I_3 = \{O_1, O_2\}$ is created at the occlusion instant. The new tracks detected after the occlusion are correctly associated to the original objects. The third example illustrates a more complex case in which there is a meeting of a couple of pedestrians, temporary occluded by a third pedestrian. Again the algorithm described in the paper manages to correctly solve this situation. It is remarked that O_5 is a compound object which includes a simple object O_4 and a compound object O_3. This leads to a hierarchy of objects similar to the dendogram representation used in clustering problems [8].

During the tracking experiments, several features are computed for each object (e.g., area, velocity, trajectory, etc). Figure 9 shows two such examples: the evolution of the object area in the first video sequence and the velocities of the objects in the third sequence.

6. Conclusion

This paper presents an algorithm to track multiple pedestrians with occlusions caused by the scene or by group creation. An hierarchical approach is used to tackle these difficulties which allows the recovery of the pedestrian tracks as soon as they become visible again. The proposed algorithm was tested with different video sequences. Tracking results are presented showing the ability of the proposed methods to perform long term tracking in the presence of complex interactions.

References

[1] A. Baumberg and D. Hogg, Learning deformable models for tracking the human body, in M. Shah and R. Jain (Ed.), *Motion-Based Recognition*, 3 (Dordrecht: Kluwer, 1997) 39-60.

[2] A. Blake and M. Isard, *Active Contours*, (Springer-Verlag, 1998).

[3] M. Ebbecke, M. Ben Hadj Ali and A. Dengel, Real time object detection, tracking and classification in monocular image sequences of road traffic scenes, *Proc. of IEEE Int. Conf. On Image Processing*, Santa Barbara, USA, 1997, Vol.2 402-405.

[4] I. Cohen and G. Medioni, Detection and tracking moving objects for video surveillance, *IEEE Proc. Comp. Vision and Pattern Recogn.*, Fort Collins, USA, 1999, .

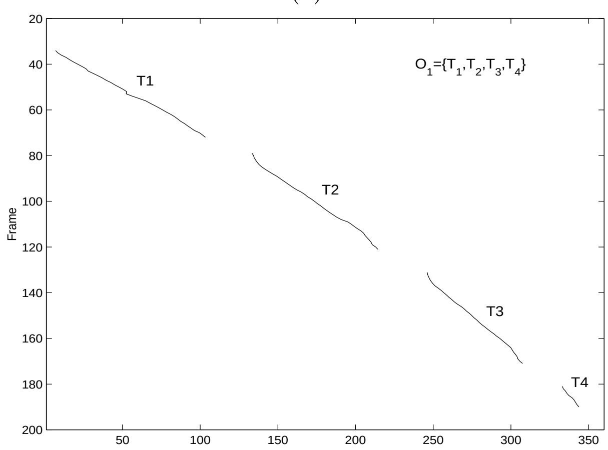
[5] I. Haritaoglu, D. Harwood and L. Davis, W^4 : Real-time surveillance of people and their activities, *IEEE Trans. On Pattern Analysis and Mach. Intell.*, 22(8), 2000, 809-830.

[6] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, Pfinder: Real-time tracking of the human body, *IEEE Trans. On Pattern Analysis and Mach. Intell*, 19(7), 1997, 780-785.

[7] C. Stauffer and W. Grimson, Leanrnig patterns of activity using real-time tracking, *IEEE Trans. On Pattern Analysis and Mach. Intell*, 22(8), 2000, 747-757.

[8] R. Duda and P. Hart, *Pattern Classification and Scene Analisys*, (New York: John Wiley and Sons Inc., 1973).

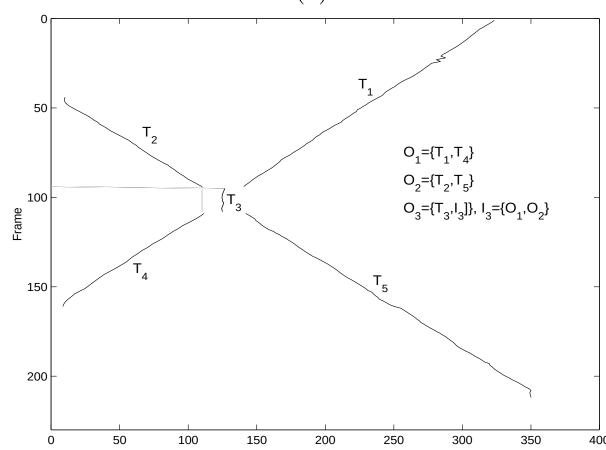
(a)



(b)

Figure 6. (a) A frame from sequence 1 and (b) the corresponding x-t plot.

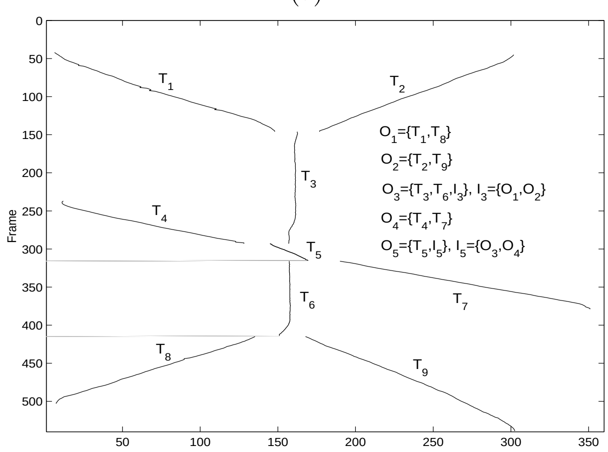
(a)



(b)

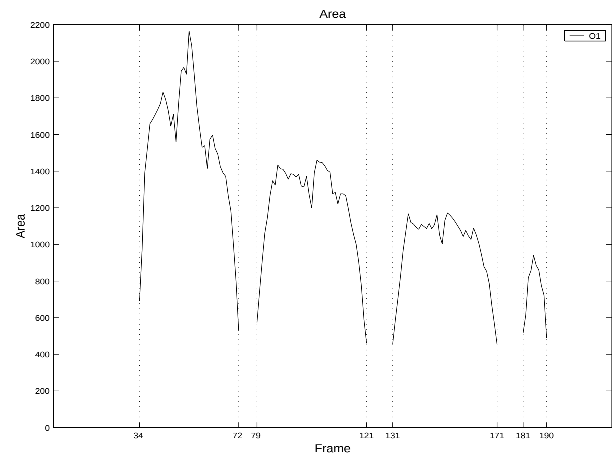
Figure 7. (a) A frame from sequence 2 and (b) the corresponding x-t plot.

(a)

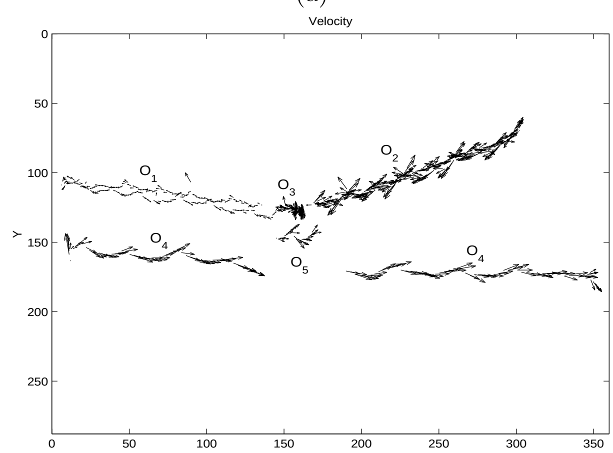


(b)

Figure 8. (a) A frame from sequence 3 and (b) the corresponding x-t plot.



(a)



(b)

Figure 9. (a) Evolution of the object area in sequence 1 and (b) evolution of the object velocities in sequence 3.