People Detection using Gait for Visual Surveillance

Imed BOUCHRIKA, Mark S. NIXON

ISIS, Department of Electronics and Computer Science
University of Southampton, SO17 1BJ, UK
{ib04r, msn}@ecs.soton.ac.uk

1 Introduction

Detecting, tracking and recognizing people using a single camera is a challenging problem due to occlusion, shadows, entry and exit of objects into the scene, and natural background clutter. Furthermore, the flexible structure of the human body, which encompasses a wide range of possible motion transformations, exacerbates difficulties for developing a vision-based surveillance system. We propose a multi-object tracking method based on feature correspondence between consecutive frames. Moving objects are assigned to different layers whereby blobs corresponding to the same object are assigned to the same layer. The criteria for allocating objects to layers is based on the Mahalanobis distance measure of shape-based features. Because of the dearth of visual surveillance systems that exploit human gait for object classification and their limited aim to detect people only using simple shape-based features extracted from silhouettes, we have explored an alternative technique for walking people detection based on their gait motion. The novelty of our approach is motivated by the latest research for people identification using gait.

2 Moving Regions Correspondence

The first problem for automated surveillance is the detection of moving objects in the scene. This is often performed via background subtraction. The adaptive background subtraction proposed by Stauffer and Grimson [**?**] is being adopted in this work. The method models detecting moving objects as temporal templates characterized with three features: size, centroid position, and aspect ratio of the bounding box. Moving objects are assigned to different layers, such that moving regions which correspond to the same object are allocated to the same layer. Each layer is defined by four parameters $L_i < s_i, a_i, x_i, y_i >$ where i is the layer index. s_i and a_i are the mean values for the sizes and aspect ratios of objects belonging to the i^{th} layer respectively. x_i and y_i are the predicted centroid position of the object in the next frame. The centroid position is estimated linearly via computing the velocity V_i as the spatial difference of the last two previous positions. We have defined a number of constraints to the allocation criteria to handle occlusion and entry and exit of moving objects into the monitored scene. A candidate will be allocated to layer L_i only if:

- The layer L_i has the smallest cost value C which is based on the Mahalanobis distance measure.
- $|s_i - S| < 3\sigma_i$ where S is the size of the candidate object, σ_i and s_i are the standard deviation and mean values of objects' sizes belonging to the i^{th} layer.
- If a candidate object does not have a corresponding layer, it will be allocated to layer L_i, if the object is mostly contained within the bounding box of L_i.

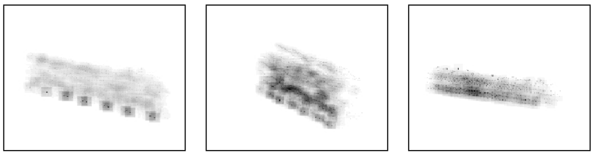
If an object is not assigned to one of the existing layers, a new layer is created for this new object. To cope with the appearance of uninteresting regions such as background clutter, we define a threshold $T = 5$, if a layer has a life span of T frames or less, then this layer is ignored and deleted.

3 People Detection

Our method classifies moving objects into either: i) person, ii) group of people or iii) undefined objects (such as vehicles). The classification procedure is based on the analysis of gait motion. During the strike phase, the foot of the striking leg stays at the same position for half a gait cycle, whilst the rest of the human body moves forward. Therefore, if we apply the corner detector on the sequence of frames, a dense area of corners will be produced at the location of the strikes. In order to locate these areas, we have estimated a measure for density of proximity. The value of proximity at point p is dependent on the number of corners within the region R_p and their corresponding distances from p. R_p is assumed to be a square area with centre p, and radius of r that is determined as the ratio of total image points to the total of corners in C_i which is about 10. We have first computed proximity value d_p of corners for all regions R_p in C_i using equation (**??**). This is an iterative process starting from a radius r. The process then iterates to accumulate proximity values of corners for point p.

$$\begin{cases} d_p^r = \frac{N_r}{r} \\ d_p^i = d_p^{i+1} + \frac{N_i}{i} \end{cases} \tag{1}$$

where d_p^i is the proximity value for rings of radius i away from the centre p, and N_i is the number of corners which are of distance i from the centre, rings are single pixel wide. Afterwards, we accumulate all the densities for the subregions R_p for all points p into one image to produce the corners proximity image. An output of the corner proximity image for different moving objects is shown in Figure (2). The corner proximity image shows brighter peaks at the heel strikes areas. A similar algorithm to [**?**] is used to derive the positions of the peaks as local maximas.



(a) (b) (c)

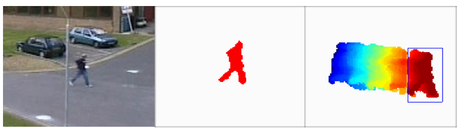
Figure 1: The Corners Proximity Images for : (a) Single Walking Person, (b) Group of People, (c) Moving Vehicle

Clearly, the corner proximity image for walking subjects has larger peaks at the bottom as legs have static periods. Furthermore, since gait is periodic, the stride length should be the same for different gait cycles, therefore the standard deviation of distances between two close strikes should tend to zero. For the classification of moving objects, we define a feature vector $< b, \sigma, \alpha >$ where b is the proportion of the lower part of the proximity image, and σ is the standard deviation value of distances between two successive peaks. α is the aspect ration of height to width of the bounding box.

4 Experimental Results

To demonstrate the efficacy of our method for automated visual surveillance, we applied on a set of videos provided by PETS 2001 compressed in JPEG format with a reduced size of 384x288. Moving objects are tracked successfully during their life span in the monitored scene. Furthermore, the system can handle occlusion efficiently, and reallocate the occluded object to the correct layer when occlusion vanishes. The appearance of uninteresting regions such as background clutter are ignored by the system as shown in Figure (2).

(a) Tracking During Occlusion



(b) Tracking recovery results after occlusion

Figure 2: Experimental Results for Handling Occlusion.

To verify the effectiveness of our approach to classify moving objects by their gait pattern, we have carried out a number of experiments on the whole PETS video data containing a total of 26 moving objects consisting of: 15 single walking subjects, 4 groups of people and 7 moving vehicles. The leave-one-out validation rule is used to evaluate the classification performance using the k-nearest neighbor classifier. The system was able to discriminate between single walking people, a group of people and vehicles efficiently using the proposed features and achieved a detection rate of %100. The feature vectors for the moving objects are projected into the feature space shown in Figure (5). Although, the classification results were promising, we have conducted further experiments to confirm the robustness of the proposed method for extracting the heel strikes, we have run the algorithm on a set of 120 video sequences from the SOTON database. The mean error for the positions of the strikes extracted by the algorithm compared to manually labelled strikes is %0.52 of the person height. The error is measured by Euclidean distance normalized to a percentage of the person's height.

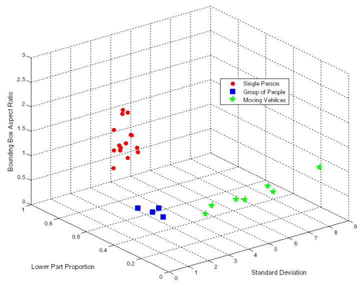


Figure 3: Feature Space

5 Conclusions

We have proposed a new method to classify moving objects for automated visual surveillance. Multiple objects are tracked successfully through the use of shape-based parameters to allocate them to different layers. Problems encountered during tracking such us background clutter, appearance of uninteresting objects and entry and exit of objects are handled efficiently. Finally moving regions are classified into either single walking person, group of people or undefined object such as vehicle. In contrast to approaches that employ shape-based parameters for classification, we have explored an alternative technique for walking people detection based their gait motion. The experimental results confirmed the robustness of our method to discriminate between single walking person, group of people and vehicle with a CCR of %100.

References

[1] H. Fujiyoshi, A. J. Lipton, and T. Kanade. Real-time human motion analysis by image skeletonization. *IEICE Trans on Information and System*, pages 113–120, 2004.

[2] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *IEEE TPAMI*, 22(8):747–757, 2000.