Using Gait Features for Improving Walking People Detection

I. Bouchrika, J. N. Carter & M. S. Nixon
Dept of Electronics & Computer Science
University of Southampton, SO17 1BJ, UK
{ib04r,jnc,msn}@ecs.soton.ac.uk

R. Mörzinger, G. Thallinger
Joanneum Research
Steyrergasse 17,8010 Graz, Austria
roland.moerzinger@joanneum.at

Abstract

In this paper, we explore a new approach for enriching the HoG method for pedestrian detection in an unconstrained outdoor environment. The proposed algorithm is based on using gait motion since the rhythmic footprint pattern for walking people is considered the stable and characteristic feature for the detection of walking people. The novelty of our approach is motivated by the latest research for people identification using gait. The experimental results confirmed the robustness of our method to enhance HoG to detect walking people as well as to discriminate between single walking subject, groups of people and vehicles with a detection rate of 100%. Furthermore, the results revealed the potential of our method to be used in visual surveillance systems for identity tracking over different camera views.

1. Introduction

In recent years, automatic visual surveillance has received considerable interest in the computer vision community. This is due to the inability of human operators to monitor the large growing numbers of cameras deployed in sensitive areas as well as the increasing concern about security and terror threats. The main aim of a surveillance system is to detect and track people in a scene, to understand their behaviour and to report any suspicious activities. The system could also recognize subject's identity by biometrics.

Detection of people is of prime importance for most monitoring application as it lays the foundation for subsequent phases of an automated system, such as identity tracking or activity recognition. However, detecting and tracking 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 vision-based surveillance system.

Existing surveillance systems for people detection are classified into several categories [6] according to their type (single or multiple camera) and their functionality (tracking single , multiple people, etc.). The W^4 [6] surveillance system employs an appearance model to track people whereby single or group are distinguished using a projection histogram. Each person in the group is located through the tracking of their head. Lipton *et al* [8] proposed a real time vision-based system to classify moving objects into either human or vehicle based on the "dispersedness". In his work, people are assumed to have a dispersedness value smaller than vehicles, however shape metrics can vary depending on image size and distance from camera. Furthermore, Viola *et al.*[11] proposed a real-time pedestrian detection system based on AdaBoost. The system employs both image intensity as well as motion information.

Wang [12] surveyed two type of features used for people detection in surveillance systems: shape-based or motion-based features. The first type relies on the shape of human silhouettes such as dispersedenss [8], aspect ratio of bounding box, or just simple shape parameters. For the motion-based features, the periodicity of human motion is considered as a strong cue for people detection. Cutler [3] described a real time method for measuring periodicity for periodic motion based on self-similarity. Javed *et al* [7] proposed a simple measurement based repeated internal motion.

Because of the dearth of visual surveillance systems that exploit human gait for pedestrian detection and their limited aim to detect people only using simple shape-based features extracted from silhouettes, we have explored a technique that improves the results obtained from an existing pedestrian detection based on the rhythmic pattern of their gait motion. The novelty of our approach is motivated by the latest research for people identification using gait [9]. In our method, people are detected through the extraction of their heel strikes, whereby stride and cadence can be estimated easily. In contrast to earlier methods, our approach does not require the subject to walk in sagittal view, as the gait pattern can be extracted from different viewpoints. This proposed method has a great potential for visual surveillance systems to incorporate a biometric system for identity tracking [2] or recognition [1].

2 Initial Person Detection and Tracking

For collecting the person detections we use the publicly available implementation of the Histograms of Oriented Gradients based pedestrian detector from Dalal *et al.* [4]. This detector achieves state-of-the-art performance on full-body pedestrian detection [5]. For each input image the detector classifies detection windows at multiple scales and locations into no pedestrian or pedestrian. In order to increase the recall of person detection in difficult conditions such as crowded scenes with cluttered background and partially occluded body parts, we introduce a simple method for person tracking. Specifically, the box of each detected person is propagated to the next frame towards the dominant direction and displacement, calculated from the mode of the orientation histogram of motion features. The motion features are selected by a sparse optical flow method based on the KLT feature tracker [10]. Features that do not match the dominant direction and displacement are discarded. The propagation step is repeated until the person is redetected in the same image region or until the number of features or covered image area between the features is below a threshold.

Figure (2) shows example detection results. Although the person detector was not able to detect the object due to partial occlusions, the person could be marked using the proposed propagation technique.

(a) Detection (b) Tracking (c) Non walking

(d) False (e) Missed (f) Misaligned

Figure 1. HoG based person detection results. Positive detection of a walking (a) and non walking human (c). Expected errors are false, missed and misaligned detections (d-f). In case of partial occlusions (b), detections are tracked using feature point motion (color coded arrows).

3 Improving Person Detection using Gait

Based on the detection provided from the person detection and tracking algorithm described earlier, we propose a validation method that can verify the presence of walking people by their gait motion within the resulting bounding boxes. The rhythmic pattern of gait motion is utilised as the main cue to distinguish walking subjects from other moving objects In order to derive a set of features for the detection of walking pedestrians, we apply frame differencing to compute a motion map image based on the change detection for the inter-frame difference. A motion map M_t at frame t is computed based on a pixel-wise variation through of a window of consecutive frames. An accumulation process is applied on the motion map by dividing the map into a grid with smaller bins of size 10x10 pixels and afterwards, summing the values in each bin. A threshold is then applied to the accumulated image. Finally, Connected Component Analysis is applied to derive the larger blobs which correspond to moving objects.

Because the striking foot is stabilised for half a gait cycle, a dense area of points is detected in the region where the foot strikes the ground. In order to locate these areas, we have devised a measure for point proximity in an image to find where the crowded region in a given image. The value of proximity at point p is dependent on the number of points within the neighbourhood region R_p and their corresponding distances from p. For simplicity, R_p is assumed to be a square area with centre p, and width $2r$ (r is set experimentally to 10 based on the image size). In order to compute the proximity image, we initially compute the neighbourhood proximity d_p for the region R_p corresponding to the point p, such that d_p is also a square region with the same width as R_p. The computation is carried out in an iterative process starting from the boundaries of R_p. It computes the nearness value of points with respect to the centre p and then it iterates inside and accumulate the previous computed values as expressed in the following equation:

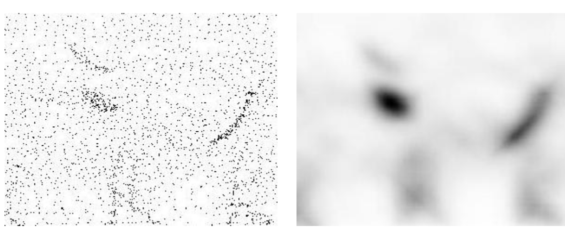
$$\begin{cases} d_p^r = \frac{N_r}{r} \\ d_p^i = d_p^{i+1} + \frac{N_i}{i} \end{cases} \quad (1)$$

where d_p^i is the proximity value for rings of distance i away from the centre p, and N_i is the number of corners at distance i from the centre.

Afterwards, in order to produce the proximity image, we accumulate all the neighbourhood proximity values d_p for all points p into one image as described in the following equation:

$$D = \sum_{x=0}^{X} \sum_{y=0}^{Y} shift(d_{(x,y)}, x, y) \quad (2)$$

where X and Y are the width and height of the image respectively. $d_{(x,y)}$ is the neighbourhood proximity value for region $R_{(x,y)}$. The $shift$ function places the proximity value $d_{(x,y)}$ on a blank image of size $X \times Y$ at the position (x, y). An output of the point proximity for a sample image is shown in Figure (2). The input image contains a point cloud with a number of dense regions. The resulting image has darker areas which correspond to the crowded regions in the input image. Because it is a challenging task to formally determine which regions in a given image are crowded or dense as opposed to using a simple 2D histogram function, the problem becomes a question of detecting darker regions of the derived proximity image.



(a) (b)

Figure 2. Example for the Proximity Measure: (a) Input Image, (b) Proximity Image

For the application context of this research, we applied the proximity measure on different moving objects being captured using a surveillance camera. Moving objects include a single walking individual, a group of two subjects walking together and a vehicle as shown in Figure (3). Clearly, the proximity image produced from the motion maps of a walking subject has a pattern of darker spots being detected at the bottom part of the image as the foot strikes the ground. Moreover, these darker regions are observed to have mostly the same level of darkness with consistent distance between two consecutive regions. On the other hand, the proximity image of people walking together constitutes a noisy pattern corresponding to the footsteps of subjects, however, the lower part of the proximity image is darker than the upper part of the image. For vehicles, the proximity image has almost a flat and consistent pattern with peaks located at random positions in the image.

Figure 3. Classification of Moving Objects

In order to derive as many features as possible that would assist with the detection of walking people from the image, gait features are derived from a sequence of regions extracted via the HoG detector. The proximity approach is then applied to subset of images where the HoG tracker has detected the presence of the subject. Fourier analysis reveals the periodic components and the 7 heighest magnitude components are used as features along with candidate extracted striking points, aspect ratio as well as the distribution of dense regions in the proximity image.

4 Experimental Results

To demonstrate the efficacy of our approach for the use in automated visual surveillance, the proposed methods for detecting walking pedestrians have been evaluated on publicly available datasets. Initially, the enhanced HoG method is applied on a selection of the PETS2006 [1] to generate bounding boxes for the possible detection of people. The gait-based approach is thereafter applied to validate the existence of pedestrian

[1] Available at : http://ftp.cs.rdg.ac.uk/PETS2006/

	HoG Detection			Gait
	Missed	True	False	CCR
PETS2006	3	5	2	100%
iLids2009	23	20	0	100%

Table 1. Pedestrian Detection Results.

through the use of gait motion. Two false detections generated by the HoG detector were filtered out successfully by the gait approach. For cases where people are missed by the HoG detector provided that they were detected in previous frames, the region correspondence method used to temporally track moving regions, was able to deduce the missed detections.

Additional evaluation was carried out on the iLids 2009 dataset provided the UK Home Office. The dataset is taken from CCTV cameras at Gatwick International Airport. The main drawback of the proposed approach is the inability to perform well in crowded scenes due to the fact the gait features are occluded. The summary of the results of the detection of pedestrians from both datasets is shown in Table (1).

To verify the effectiveness of our approach to classify moving objects using gait, we have carried out experiments on the PETS2001 video data containing a total of 27 moving objects being annotated from the set of four video sequences. The leave-one-out validation rule is used to assess the performance of the classification using the k-nearest neighbour classifier. The system is able to discriminate between a single walking subject, a group of people and vehicles efficiently using the proposed features, achieving a Correct Detection Rate of 100%. The feature vectors of moving objects are projected into the feature space shown in Figure (4). Clearly that the gait pattern is a strong cue to distinguish between walking people and vehicles. This is consistent with the findings of BenAbdelkader *et al.* [1] where the stride parameters were utilised for people identification.

5 Conclusions

We have proposed a new approach for enriching the HoG method for pedestrian detection for automated visual surveillance. In contrast to approaches that employ shape-based parameters for classification, we have explored an alternative technique for walking people detection based the rhythmic pattern of their gait motion. Initial evaluations of the method on two different datasets showed that gait can be considered a highly descrimitive feature for detecting pedestrians. The experimental results confirmed the robustness of our method to discriminate between single walking person, group of people and vehicle with a classification rate of %100. The proposed method confirmed its potency to be extended to tracking people identity based their cadence and stride across different camera views and we are to investigate this further on different datasets.

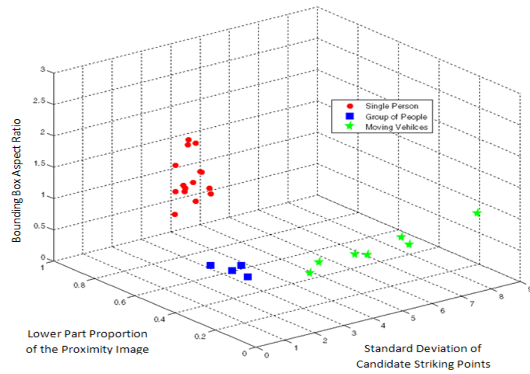


Figure 4. Classification Feature Space

References

[1] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and Cadence as a Biometric in Automatic Person Identification and Verification. *IEEE Intl Conf on AFGR*, pages 357–362, 2002.

[2] I. Bouchrika, J. Carter, and M. Nixon. Recognizing People in Non-Intersecting Camera Views. *Intl Conf on ICDP*, 2009.

[3] R. Cutler and L. S. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE TPAMI*, 22(8), 2003.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Conf on CVPR*, 2005.

[5] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *Conf on CVPR*, 2009.

[6] I. Haritaoglu, D. Harwood, and L. S. Davis. W4: real-time surveillance of people and their activities. *IEEE TPAMI*, 22(8), 2000.

[7] O. Javed and M. Shah. Tracking and object classification for automated surveillance. *In Proc ECCV*, 4, 2002.

[8] A. J. Lipton, H. Fujiyoshi, and R. S. Patil. Moving target classification and tracking from real-time video. *IEEE Workshop on Application of CV*, 1998.

[9] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2005.

[10] C. Tomasi and J. Shi. Good features to track. In *IEEE Conf on CVPR*, pages 593–600, 1994.

[11] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision*, 63(2):153–161, 2005.

[12] L. Wang, W. Hu, and T. Tan. Recent Developments in Human Motion Analysis. *Pattern Recognition*, 36(3):585–601, 2003.