10-5

MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN

Detection of Approaching Pedestrians from a Distance Using Temporal Intensity Patterns

Siu-Ming Cheung
Dept. of Computer Science & Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
smcheung@cse.cuhk.edu.hk

Yiu-Sang Moon
Dept. of Computer Science & Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
ysmoon@cse.cuhk.edu.hk

Abstract

Pedestrians captured from real world surveillance cameras can often be in frontal view. This is true for surveillance cameras installed in bridges and corridors. Surveillance cameras in these environments are often oriented in the direction along the passage. In such a setting, many of the walking pedestrians would often appear to be approaching the camera. In this paper, we focus on the detection of these pedestrians from a distance. Common pedestrian detection algorithms such as background subtraction or motion detection do not work well for approaching pedestrians in the distance when viewing the pedestrians from the front view. We, therefore, propose to detect such pedestrians using temporal pattern of intensity observed from a pair of human walking legs. Thorough discussion of the detection algorithm is presented in this paper. Evaluation results show that pedestrians of image size as small as 30 pixels tall can be detected using our algorithm. Our method can be utilized in intelligent surveillance systems for initiating tracking process and subsequent recognition of pedestrians from a distance.

1 Introduction

With the massive deployment of surveillance cameras, there is increasing interest on the semantic understanding of the videos taken by surveillance cameras. Among the technologies in semantic understanding of surveillance video, pedestrian detection is perhaps the most important one. This is because a number of automatic surveillance tasks heavily rely on the success of pedestrian detection. Examples include recognition of pedestrians inside a designated area [12][3], and monitoring of their activity [1].

Pedestrians captured from real world surveillance cameras can often be in frontal view. This is true for surveillance cameras installed in long passages such as corridors, bridges, and streets. Surveillance cameras in these environments are often installed with an orientation facing the pedestrians walking along the passage. In such a setting, pedestrians would often appear to be either frontal or rear. In this paper, we are particularly interested in detecting pedestrians in frontal view. We would refer them as approaching pedestrians. In our work, we assume approaching pedestrians are frontal to a tolerance of $\pm 15°$ in the view angles. Among the approaching pedestrians in a scene, we focus on the detection of those from a distance because *a*) such approaching pedestrians can be detected, tracked, and analyzed earlier, *b*) face recognition is feasible because of smaller elevation angle . As our task is to detect distant pedestrians, we would refine the definition of them as those walking people at a distance of at least 20m away from a common web camera so that they appear to be 30 to 50 pixels tall.

The most commonly used methods to detect pedestrians in surveillance systems are the background subtraction techniques [11]. However, the output of these methods is usually noisy, with small blobs of false objects. It is difficult to differentiate between false objects and distant pedestrians. Another issue is that distant approaching pedestrians can be absorbed into the background model, as they appear to be motionless and stay at roughly the same image location over time [8]. An alternative approach to detect pedestrians is by motion detection. An implicit assumption of this approach is that the image locations of pedestrians form significant tracks in a short period of time. This assumption is valid if pedestrians appear in side view or aerial view. For images of approaching pedestrians taken at the front view, however, the tracks are often insignificant. Some researchers use shape or gradient features to detection pedestrians. Representative works in this area are given by [5] and [7]. However, these methods in general require a specific minimum image size of pedestrians, and are therefore not suitable for detection of distant pedestrians.

Our approach for the detection of distant approaching pedestrians is motivated by two observations. The first is that human usually walk in a characteristic way [4], i.e. repeatedly raise legs and step forward. This property of human walking has been utilized in many motion-based pedestrian detection algorithms [6][9]. The second one is that pedestrians are usually illuminated from the top of them, regardless of whether the environment is indoor or outdoor. Based on these observations, we believe that distant approaching pedestrians can be detected by the temporal intensity pattern observed on a pair of walking human legs. We assume this temporal intensity pattern is a distinctive feature of walking human. Our proposed method works by extracting temporal patterns of intensity from different image locations and cross-correlate them with templates of off-line generated patterns which correspond to walking human. We have evaluated our method using video sequences and obtain satisfactory results. The contents of this paper is organized as follows. In section 2, the proposed method is presented. Section 3 demonstrates the result on a testing video. Section 4 gives the conclusion of our work.

2 Proposed Method

Our detection algorithm works on a temporal window of 1 to 2 seconds. To detect distant pedestrians from a given a video sequence, our algorithm first divided it into short sequences of video frames. Each of these short sequences should be about 1 to 2 seconds in length. The purpose of this preprocessing step is to ensure complete human gait cycles are captured, while at the same time the spatial locations and scales of pedestrians are kept roughly the same within the short video sequences. To detect pedestrians form each of these video sequences, we apply sliding-window method to exhaustively 'scan' over spatial locations and scales: For each image location, temporal intensity patterns are extracted from rectangular windows on the image plane. We refer these windows as detection window. If prior knowledge about scene geometry is available, the detection window can be limited to image locations where presence of distant pedestrians is expected. The intensity pattern collected from a detection window is matched against off-line generated templates. If the temporal intensity pattern extracted is closed enough to any of the templates, the algorithm would report a detected pedestrian at that spatial location.

2.1 Tracking of detection window

It should be noted that although our algorithm handles only a short sequence of video frames each time, slight displacement of pedestrians within the video frames is inevitable. To prevent the subsequent feature collection process from being affected by this slight displacement, we calculate the motion vector of the detection window, in each video frames. The motion vector can be found using optical flow estimation methods, such as the Pyramidal LKT tracker [2]. In our work, we assume the scales of distant pedestrians do not change significantly with time, this assumption is reasonable for distant pedestrians.

2.2 Temporal intensity pattern

For each of the detection windows, classification is performed to determine whether the window is a tight bounding box of a pedestrian. Features of intensity pattern are extracted from four rectangular regions in a detection window (see Figure 1).

The locations of these regions are specified such that if the detection window is a tight bounding box of a pedestrian figure, the rectangular regions would each cover a human leg segment. From each of these regions, we calculate the spatial mean intensities for all video frame and denote them as $r_k(t)$, where k and t are respectively the region and frame index. The spatial mean can be efficiently computed using integral image technique. Since pedestrians can wear any kind of clothing, the magnitude and variation of intensity pattern vary with people. In our application, since our concern is to detect the temporal intensity variation of leg segment, therefore the spatial intensity values $r_k(t)$ are normalized with respect to time:

$$\hat{r}_k(t) = \frac{r_k(t) - \mu_k}{\sigma_k}$$

where μ_k and σ_k are the temporal mean and standard deviation of $r_k(t)$. Examples of normalized intensity values $\hat{r}_k(t)$ are plotted in Figure 2. We can observed that the intensity patterns from detection windows that contain pedestrians and those that do not contain pedestrian are quite different. This suggests that the temporal intensity pattern can be used to differentiate between pedestrians and non-pedestrians. The intensity pattern observed from pedestrians can be explained by the difference in reflected light intensity because of the change in leg segment angles.

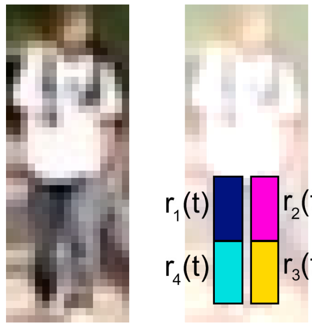


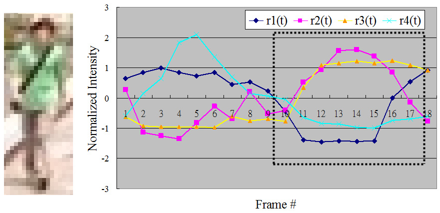
Figure 1: The figure shows a detection window at the location of a distant pedestrian. The interior rectangles labeled with $r_k(t)$ show the spatial locations of the four regions in the detection window, each corresponds to a leg segment. Spatial mean intensity values $r_k(t)$ are estimated from these regions.

2.3 Classification of intensity pattern

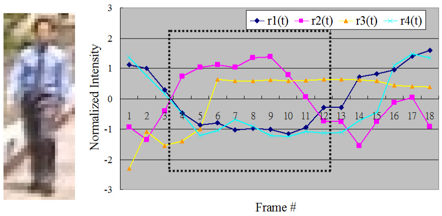
The aim of the classification step to decide whether a real pedestrian figure exists in a detection window, based on the temporal intensity pattern extracted from the detection window. In our proposed method, classification is done by matching the pattern with offline generated templates. We have built two templates, corresponding to walking pedestrians with different walking speeds. These templates are built by acquiring the joint angles of walking pedestrians and then simulating the temporal intensity pattern of leg segments using Phong reflection model [10]. It should be noted that there is usually time lag between the pattern extracted from pedestrians and that of the templates, i.e. the patterns belongs to walking pedestrians with different initial walking pose. The extracted temporal intensity pattern is matched with template using cross-correlation. Cross-correlation is particular suitable for our application because it define a similarity function in terms of the time lag between two signals. The similarity function is defined as:

$$S(\theta) = \sum_{t=1}^{T} \sum_{k=1}^{4} \hat{r}_k(t) p_k(t+\theta)$$

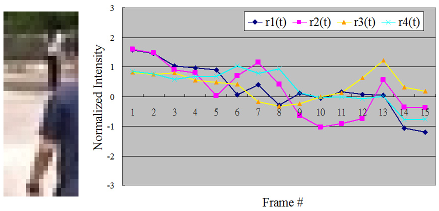
where θ is the time lag between the pattern of observed intensity and the intensity in template, T is the size of the temporal window, and $p_k(\cdot)$ is the intensity pattern of the template. Our algorithm would report a detection if the value of the similarity function exceed a tunable threshold τ for any time lag θ.



(a)



(b)



(c)

Figure 2: (a), (b): Normalized intensity features extracted from detection windows of pedestrians. The patterns are similar. The rectangles with dotted lines show how they can be match with each other. (c): The normalized intensity pattern extracted from a detection window of non-pedestrian, which is apparently different from those of (a) and (b).

3 Experimental Results

The purposes of our experiments are to evaluate the proposed detection algorithm in terms of the following: *1*) the range of apparent sizes of pedestrians that can be detected, and *2*) detection rate of approaching pedestrians of 30 to 50 pixels tall . To evaluate the above, we have collected a video from a long bridge. In this video, pedestrians were walking either towards or away from the camera. The resolution of the video is 800×600. Pedestrians at the far end of the bridge appear to be 15×30, while those in the midway of the bridge appear to be 30×60. Therefore, distance pedestrians in this testing video are those along the midway to the far end of the bridge. Since our algorithm focus on the detection of distant and frontal pedestrians, the detection rate in our experiments is defined as the number of detected frontal and distant pedestrians divided by the total number of frontal and distant pedestrians.

We extract segments of short video frames, in which there are approaching pedestrians at the far end of the bridge. The video segments are passed to our algorithm for detection, and detection windows in 8 different sizes from 15×30 to 50×100 are used. The results are shown in Figure 3. In the video segments, pedestrians at the far end (Figure 3c, 3h) to the midway (Figure 3a, 3f, 3l) of the bridge can be detected. Pedestrians near the camera (Figure 3b, 3k) are ignored by our detection algorithm, as we only aim at detecting distant pedestrians. In these segments, there are 14 approaching pedestrians (3 of them are the same person). Our algorithm is able to identify 12 of them. The missed pedestrians are shown in Figure 3b and 3e. The reason of missed detection in Figure 3b is that the pedestrian is running. In Figure 3e, the reason of missed detection is that the shape of the pant of the pedestrian deforms quickly, such that it does not appear to be rectangular block, as in Figure 1. Note that although there are 8 false positive detections, 7 of them are pedestrians from rear view. It suggests that more work is needed to differentiate between pedestrians in frontal and rear view. A possible solution is to use tracking result to prune those pedestrians in rear view. The false detection in Figure 3h is due to the motion of a waving banner. In our limited testing data, the recall rate is 85.7%. We think this result is satisfactory, as pedestrians from a distance are more difficult to detect than those near the camera.

4 Conclusion

Our key contribution is the proposal of using temporal intensity pattern for the detection of approaching pedestrians from a distance using their front view images. We have presented a detection algorithm based on the proposed intensity pattern. The algorithm can detect multiple pedestrians at the same time. We evaluated the detection algorithms using video of a long bridge, and the result showed that satisfactory results can be obtained even if only a few standard templates are used. We expect the detection results will be useful to guide surveillance system to alert for approaching pedestrians, and perform analysis on them. Future work will include estimation of the distance of pedestrians from camera and tracking of their location.

Acknowledgements

The work described in this paper was substantially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 415207).

References

[1] J. Ben-Arie, Z. Wang, P. Pandit, S. Rajaram: "Human Activity Recognition Using Multidimensional Indexing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.8, 2002.

[2] J. Y. Bouguet: "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm," Technical report, Intel Corporation Research Labs, 2002.

[3] F. Chan, Y. S. Moon, J. S. Chen, Y. K. Ma, W. H. Tsang, K. K. Fu: "A Unified Framework for Capturing Facial Images in Video Surveillance Systems Using Cooperative Camera System," *SPIE Symposium on Defense & Security*, vol.6971, 2008.

[4] R. Cutler, L. S. Davis: "Robust Real-Time Periodic Motion Detection, Analysis, and Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, no.8, 2000.

Figure 3: (a)-(l): Detection result of our proposed algorithm on a video sequence taken at a long bridge. Detection windows that are classified as pedestrians are represented using white rectangles. The embedded image on the top right hand corner on each figure is a close-up of the detection results. The images at lower right hand corner are close-ups of original video frames.

[5] N. Dalal, B. Triggs: "Histograms of Oriented Gradients for Human Detection," *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[6] N. Dalal, B. Triggs, C. Schmid: "Human Detection Using Oriented Histograms of Flow and Appearance," *European Conference on Computer Vision*, 2006.

[7] B. Leibe, E. Seemann, B. Schiele: "Pedestrian Detection in Crowded Scenes," *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[8] L. Li, W. Huang, I. Y. H. Gu, Q. Tian: "Foreground Object Detection from Videos Containing Complex Background," *ACM Multimedia*, 2003.

[9] T. Nakada, S. Kagami, H. Mizoguchi: "Pedestrian Detection using 3D Optical Flow Sequences for a Mobile Robot," *IEEE Sensors*, 2008.

[10] B. T. Phong: "Illumination for Computer Generated Images," *Communications of the ACM*, 18(6), 1975.

[11] M. Piccardi: "Background subtraction techniques: a review," *International Conference on System, Man and Cybernetics*, 2004.

[12] X. H. Zhou, R. T. Collins, T. Kanade, P. Metes: "A Master-Slave System to Acquire Biometric Imagery of Humans at Distance," *ACM International Workshop on Video Surveillance*, pp.113-V120, 2003.