

Combining Multiple Cues for Pedestrian Detection in Crowded Situations

Shih-Shinh Huang¹, Feng-Chia Chang¹, Ching-Hu Lu²

¹Dept. of Computer and Communication Engineering, National Kaohsiung First University of Science and Technology; ²Dept. of Information Communication, Yuan Ze University.
Email: powwhuang@gmail.com

Received April, 2013.

ABSTRACT

This paper proposes a vision-based pedestrian detection in crowded situations based on a single camera. The main idea behind our work is to fuse multiple cues so that the major challenges, such as occlusion and complex background facing in the topic of crowd detection can be successfully overcome. Based on the assumption that human heads are visible, circle Hough transform (CHT) is applied to detect all circular regions and each of which is considered as the head candidate of a pedestrian. After that, the false candidates resulting from complex background are firstly removed by using template matching algorithm. Two proposed cues called head foreground contrast (HFC) and block color relation (BCR) are incorporated for further verification. The rectangular region of every detected human is determined by the geometric relationships as well as foreground mask extracted through background subtraction process. Three videos are used to validate the proposed approach and the experimental results show that the proposed method effectively lowers the false positives at the expense of little detection rate.

Keywords: Pedestrian Detection; Circular Hough Transform; Head Foreground Contrast; Block Color Relation

1. Introduction

As wide deployment of cameras in public environment, such as airports, parting lots, and mass-transit stations, accurate estimating the number of people and locating each individual is an important issue in automation of video surveillance system. It can provide valuable information for safety control, urban planning, or business managing. However, the successful pedestrian detection in crowded situations is still a challenging problem due to severe occlusion, dynamic background and foreground clutter. The approaches for pedestrian detection can be generally divided into categories: approaches using global features and those using local features. A normal way to use global feature is based on template matching, which constructs the human templates from different viewing angles and views. It detects human by comparing the extracted shape with the constructed templates [9, 10]. This kind of approaches can avoid effects caused by complex background and noises. However, the approaches which only take the global features for human description have a tendency to fail in detecting partially occluded humans.

In the literature, the use of complementary local features which denote the partial appearance of human in the image is motivated for solving the occlusion. As we known, the Histograms of Oriented Gradients (HOGs)

are firstly proposed by Dalal *et al.* [11] and have been widely used in human detection. The improvements that make HOGs more representative were proposed in [12, 13]. Although the methods using local features effectively tackle with occlusion effect, they suffer from the problem of false detection in case of complex background or severe occlusion.

Based on the assumptions that the camera is stationary and the pedestrians are in upright standing, this paper proposes a method for detecting pedestrians in crowded situations by combining multiple features. The remainder of this paper is organized as follows. In section 2, we introduce how to generate a set of pedestrian candidates. Then, three features including shape, head foreground contrast (HFC) and block color relation (BCR) for verifying the detected candidates are discussed in Section 3. Finally, we validate the proposed method by using three videos and give some discussion.

2. Candidate Generation

In this section, we will introduce how to generate a set of possible pedestrian candidates. The foreground regions are firstly segmented by background subtraction algorithm based on an assumption that the camera is stationary. Then, a circle detection algorithm based on

Hough transform (CHT) is applied to detect all circular regions and each of which is considered as the head candidate of a pedestrian.

2.1. Foreground Segmentation

The intensity distribution of each pixel is modeled by a Gaussian distribution $N(\mu, \Sigma)$, where μ and Σ are the mean vector and correlation matrix, respectively. The first N_B frames $\{x_i\}_{i=1}^{N_B}$ are used to initialize μ and Σ as follows.

$$\mu = \frac{1}{N_B} \sum_{i=0}^{N_B} x_i \quad \Sigma = \frac{1}{N_B} \sum_{i=0}^{N_B} (x_i - \mu)(x_i - \mu)^T \quad (1)$$

In order to adapt to the background change, the parameters, μ and Σ , are updated over time and the strategy used is similar to the method proposed in [1]. For the currently observed image I , the pixel p is considered as the foreground if the probability

$\Pr(I(p) | N(p))$ is below the pre-defined threshold θ_1 , where $N(p)$ is the constructed Gaussian distribution of the pixel p . The parameter θ_1 is set to 0.1 in our work. The foreground mask $FM(\cdot)$ can be expressed as:

$$FM(p) = \begin{cases} F & \text{if } \Pr(I(p) | N(p)) < \theta_1 \\ B & \text{otherwise} \end{cases} \quad (2)$$

2.2. Circular Hough Transform

In crowd scene with severe occlusion, the head contour is a more robust cue which has the properties of low variance in appearance and high visibility from different views. Accordingly, the detection of head contour is generally served as the first step to generate a set of possible pedestrian candidates [2, 3]. In their approaches, the Ω -shape template is applied to locate the head-shoulder candidates in the image. In case of mutual occlusion, the used Ω -shape template may result in miss detection. Instead of using Ω -shape template, this work applies a circular Hough transform algorithm [4] to find out the head candidates in the edge map of the segmented foreground. The red circles shown in the left-top image of Figure x are the detected head candidates. Apparently, all heads are successfully detected but with several false alarms.

3. Candidate Verification

In this section, the shape feature and the proposed two features, head foreground contrast (HFC) and block color relation (BCR) are used to verify the detected head candidates. These features for verification are applied in a cascaded manner and the flow chart of is shown in **Figure 1**.

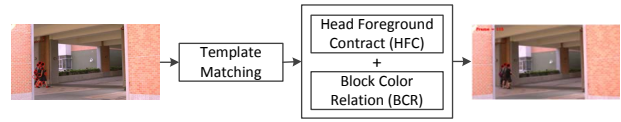


Figure 1. The flow chart of the candidate verification.

3.1. Template Matching

Pedestrian shape has been proven its robustness in describing the pedestrian appearance. The common way to utilize the shape for checking the existence of the pedestrian is by comparing it with a set of constructed templates [5]. The method to construct templates for matching is the same as our previous work [6]. Let $\{T_i\}_{i=1}^{N_T}$ be the set of the constructed templates. The images on the right-bottom in **Figure 2** are the learned templates. After obtaining edge template, the similarity of the original image I_{origin} between templates T can be calculated through the using of Chamfer distance [7] which is defined as:

$$D_{chamfer}(I_{origin}, T) = \frac{1}{|T|} \sum_{t \in T} d_{DT}(I_{origin}, t) \quad (3)$$

where $|T|$ is the number of edge points in template T and $d_{DT}(\cdot)$ is the distance transform of a specific image.

The pedestrian existence of a window determined by the circle size of the detected head candidate is firstly verified by template matching. If the minimum distance of the pedestrian window to all templates is less than a defined threshold θ_2 , this head candidate is considered as the true positive so far. θ_2 is set to 10. **Figure 2** shows the verification process of the template matching.

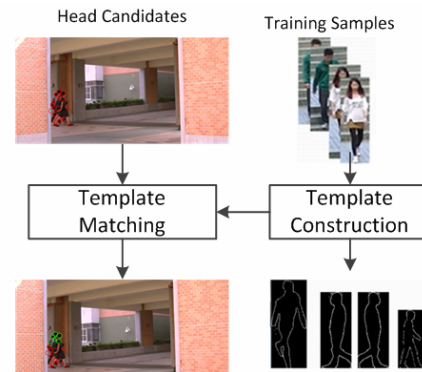


Figure 2. Verification using template matching.

3.2. Head Foreground Contrast (HFC)

When the moving direction of the pedestrians is horizontal with respect to the image, the foreground regions inside and above head have high contrast. This phenomenon is used to further eliminate the false positives. Let R_{Int} and R_{Avt} be the two regions inside and

about the detected circular head candidate, respectively. The pictures on the right hand side of **Figure 3** is the schematic description of R_{Int} and R_{Avt} . We define the existence confidence of a head candidate H using HFC as:

$$Conf_{HFC}(H) = \sum_{p \in R_{Int}} 1_{FM(x=F)}(p) - \sum_{p \in R_{Avt}} 1_{FM(x=B)}(p) \quad (4)$$

where $1_{FM(x=F)}(p)$ is an indicator function to check if the point p is a foreground one.

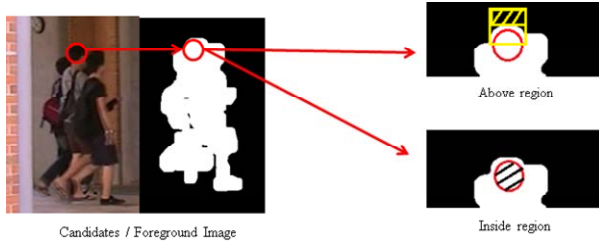


Figure 3. Region description of R_{Int} and R_{Avt} .

3.3. Block Color Relation (BCR)

According to the observation of color consistency in [2], the colors of blocks below and above head block (background) should be different and the colors of the next two blocks below head block should be similar as shown in **Figure 4**. In this work, the Bhattacharyya distance [8] is adopted to model the color similarity of two blocks. Let μ_i and Σ_i be the color mean and covariance matrix over a region R_i . Then, the color difference between two regions using the Bhattacharyya distance is defined as:

$$d_b(R_i, R_j) = \frac{1}{8} (\mu_i - \mu_j)^T \left[\frac{\Sigma_i + \Sigma_j}{2} \right]^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \frac{|\Sigma_i + \Sigma_j|}{2 \sqrt{|\Sigma_i| |\Sigma_j|}} \quad (5)$$

Consequently, the existence confidence of H using BCR is defined as:

$$Conf_{BCR}(H) = d_b(R_1, R_2) + d_b(R_3, R_4) \quad (6)$$

Then, the HCF and BCR are fused by linear weighting as (6) and the head candidates have high combined confidence is finally considered as a pedestrian head.

$$Conf(H) = 0.5 \times Conf_{HFC}(H) + 0.5 \times Conf_{BCR}(H) \quad (7)$$

The pedestrian region is determined using the circle radius and aspect ratio of human body and an example is shown in **Figure 5**.

4. Experiment

The proposed method is implemented on a platform with 2.4 GHz Intel i7 Core and 4G RAM. The OpenCV library is used to facilitate the development of our system.

In this section, we introduce the scenarios and camera setting for collecting three videos and also give the quantitative analysis of our proposed method.

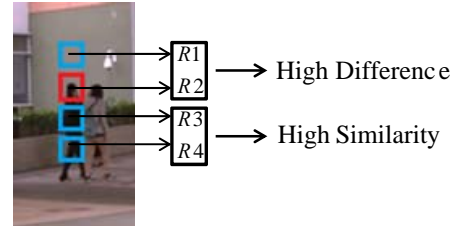


Figure 4. Block color relation. For a correct candidate, the regions R_1 and R_2 should have high color difference; the regions R_3 and R_4 should have high color similarity.

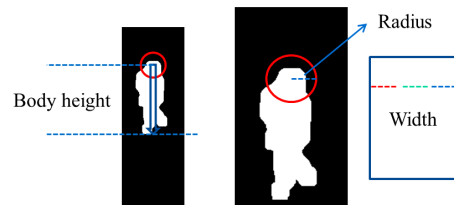


Figure 5. The determination of pedestrian region using the radius of head circle.

In our campus, three videos are collected for validating our proposed method. Each collected video has 120 frames with resolution 1280×720 and the detailed description of these three videos in terms of depression angle, moving direction, near/far, and scene are listed in **Table 1**. The first 20 frames for each video are used for background modeling and the ground truth of the remaining 100 frames are manually annotated. The criteria used for evaluating the detected results are detection rate (DR) and false alarm rate (FAR)

Table 1. Description of three collected videos.

Factor \ Video	Video1	Video2	Video3
Depression Angle	0	45	45
Moving Direction	Horizontal	Vertical	Horizontal
Near/Far	Near	Near	Far
Scene	Passage	Staircase	Square

In this section, we compare the performance of three methods and analyze the complementary property of HFC and BCR features. They are methods based on head-foreground contrast (HFC), block color relation (BCR), and both. **Table 2** shows the experimental results of these three videos and the some results are shown in **Figure 6**. Apparently, the method combining HFC and BCR will significantly reduce the FAR at the expense of

little DR. To further discuss the complementary property of HFC and BCR, we exhibit the detection rate of these three methods for every frame in **Figure 7**. Obviously, HFC and BCR are complementary of each other as the red and blue lines shown.

Table 2. Detection rate (DR) and false alarm rate (FAR) of the three approaches for three videos, respectively.

	Video1			Video2			Video3		
	HFC	BCR	Our	HFC	BCR	Our	HFC	BCR	Our
DR (%)	84	83	82	67	66	81	80	72	74
FAR(%)	15	9	7	29	39	32	31	28	24



Figure 6. Examples of the detection results in crowded situations.

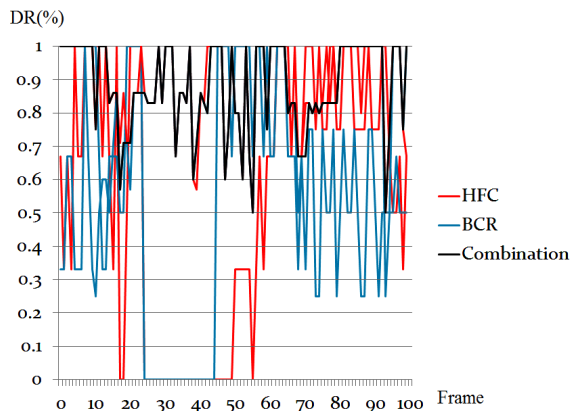


Figure 7. Complementary Analysis of HFC and BCR for the video 2.

5. Conclusions

This paper presents a method for pedestrian detection in a crowd scene. Firstly, the foreground regions are segmented from background by using background subtraction technique and the circular Hough transform is used to extract the head candidates. By combining two complementary features HFC and BCR, the experiment results of three videos show that the proposed method can reduce the false positive rate at the expense of little detection rate. However, the accuracy of the proposed method is still not yet ready for real applications. The motion cue as well as tracking strategy should be incorporated in the system in the near future.

6. Acknowledgements

This research is partially supported by the project grant 101-2221-E-327-038-.

REFERENCES

- [1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, Vol. 2, 2009, pp. 246-252.
- [2] P. Tu, et al., "Unified Crowd Segmentation," *European Conf. on Computer Vision*, 2008, pp. 691-704.
- [3] T. Zhao, "Bayesian Human Segmentation in Crowded Situations," *IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, Vol. 2, 2003, pp. 459-466.
- [4] M. Perreira Da Silva, V. Courboulay, and A. Prigent, P. Estraillier, "Fast, Low Resource, Head Detection and Tracking for Interactive Applications," *PsychNology Journal*, 2009, pp. 243-264.
- [5] D. M. Gavrila, "A Bayesian, Exemplar-Based Approach to Hierarchical Sha
- [6] S. S. Huang, C. Y. Mao, P. Y. Hsiao, and L. A. Yen, "Global Template Matching for Guiding the Learning of Human Detector," *IEEE Conf. on Systems, Man, and Cybernetics*, 2012, pp.565-570.
- [7] T. Nguyen, D. pe Matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 29, N. 8, 2007, pp. 1408-1421. P. Ogunbona, and W. Li, "Human Detection Based on Weighted Template Matching," *IEEE Intl. on Multimedia and Expo*, 2009, pp. 634-637.
- [8] S. S. Huang, L. C. Fu, and P. Y. Hsiao, "Region-Level Motion-Based Background Modeling and Subtraction Using MRFs," *IEEE Transactions on Image Processing*, Vol. 16, No. 5, 2007, pp. 1446-1456. [doi:10.1109/TIP.2007.894246](https://doi.org/10.1109/TIP.2007.894246)
- [9] S. Belongie and J. Malik, "Matching with Shape Contexts," *IEEE Workshop on Content-based Access of Image and Video Libraries*, 2000, pp. 20-26. [doi:10.1109/IVL.2000.853834](https://doi.org/10.1109/IVL.2000.853834)
- [10] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shaped-Based Pedestrian Detection," *IEEE Intelligent Vehicle Symposium*, 2000, pp. 215-220.
- [11] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, Vol. 1, 2005, pp. 886-893.
- [12] Z. Hao, B. Wang, and J. Teng, "Fast Pedestrian Detection Based on Adaboost and Probability Template Matching," *Intl. Conf. on Advanced Computer Control*, Vol. 2, 2010, pp. 390-394.
- [13] S. Paisitkriangkrai, C. Shen, and J. Zhang, "Performance Evaluation of Local Features in Human Classification and Detection," *IET Computer Vision*, Vol. 2, No. 4, 2008, pp. 236-246. [doi:10.1049/iet-cvi:20080026](https://doi.org/10.1049/iet-cvi:20080026)