# Multi-feature Detection for Pedestrian Tracking in Traffic Surveillance

Wang Xuan
School of Information Engineering
Chang'an University
Xi'an, China
jessica036@126.com

Song Huansheng
School of Information Engineering
Chang'an University
Xi'an, China
hshsong@chd.edu.cn

Cui Hua
School of Information Engineering
Chang'an University
Xi'an, China
huacui@chd.edu.cn

***Abstract*—Pedestrian tracking is an active research area to improve traffic safety for intelligent video surveillance. This paper proposes an efficient method to automatically detect and track far-away pedestrians in surveillance video using the motion feature extraction and analysis. Firstly, pedestrian features of each frame are extracted by object segmentation, recognition and feature extraction. Then, the similar features in current frame image of all candidate objects are matched by the characteristic information of pedestrians in the previous frame which is considered as a template. Finally, pedestrian trajectory analysis algorithms are used on the track trajectories and the motion information can be attained, which can realize the early classification warning of pedestrian events. Experimental results in practical surveillance demonstrate that this method shorten the processing time of matching pedestrians and improve the reliability and real-time ability of pedestrian tracking.**

***Keywords-pedestrian tracking; video surveillance; feature matching; trajectory analysis***

## I. INTRODUCTION

Pedestrian tracking is an important issue in video surveillance, which has attracted ubiquitous attention of researchers in the last decades[1-3]. It requires preprocessing which consists of the recognition of the pedestrian target in video sequences and distinction of pedestrians from other video objects. Pedestrian tracking locates the position of each detected target in the frames of analyzed video sequences. In recent years, loads of methods for pedestrian tracking including active contour tracking[4], model tracking[5], feature tracking[6], etc, have been proposed by vision researchers.

Features such as Gabor, color, edge, centroid, area and corner have a good performance in real-time tracking with less computation. In [7], JongSeok Lim and WookHyun Kim tracked pedestrians with possibly partial occlusions with block matching method using color information.Tudor Barbu [8] used a robust pedestrian tracking method based on template matching process of Histogram of Oriented Gradient (HOG). Dinh T.H. et al. [9] presented a high throughput FPGA architecture for detecting corner features on traffic images and Mohanna et al.[10]used a multi-scale corner detector and three-frame monitoring to achieve tracking. In this paper, a method based on multi-feature tracking is used to detect and track pedestrians in traffic video sequences. We combine multi-feature together and construct a mathematical model of matching function to detect pedestrians target. Furthermore, the pedestrian trajectories are analyzed by a specific method based on the motion information of pedestrians in order to confirm targets.

The rest of the paper is organized as follows: Section Ⅱ and section Ⅲ describe the method of multi-feature matching and the implementation of tracking; In section Ⅳ, the motion information of pedestrian trajectory are analyzed. Experiment results are given in section Ⅴ; The final section draws the conclusion of this paper.

## II. MULTI-FEATURE MATCHING

The pedestrian target of two adjacent frames can be matched with certain feature information. In this paper, we establish the matching function as

$$V(x) = w_1 * f_1(x) + w_2 * f_2(x) + ... + w_n * f_n(x) \quad , \tag{1}$$

where $f_1(x), f_2(x)...f_n(x)$ are the input values, $w_1, w_2...w_3$ are the corresponding weights, and $V(x)$ is the result of the function. It is obvious that input value is a decisive parameter of the matching function. Weight coefficients represent the importance of each decisive parameter. The result of the matching function is used to measure the similarity between targets. In this paper, there are three input values including location closeness, area similarity and feature matching degree. The weight coefficients are obtained by machine learning.

### *A. Centroid Position*

According to research statistics, the speed of pedestrian is generally 1.4 m/s and the shooting view of the traffic surveillance cameras is usually far. Thus we assume that the speed of pedestrians motion is uniform. What is more, the altitude variation of pedestrians is approximately linear, the centroid position is used to describe the similarity of pedestrian targets. We establish the model of pedestrian motion as

$$M_t = M_{t-1} + \Delta M \tag{2}$$

$$\Delta M = H_t V_t + \varepsilon_t \tag{3}$$

where $M_t = [X_t, Y_t]^T$ is the center-of-mass coordinate in frame $t$ . $\Delta M = [\Delta X, \Delta Y]^T$ is the position displacement of pedestrian target from frame $t$-1 to frame $t$ in the image. $H_t = \Delta t * E$ is the variation of time and $V_t = [V_x, V_y]^T$ is instantaneous speed of pedestrian targets. $\varepsilon_t = [\varepsilon_x, \varepsilon_y]$ is the displacement error.

Analyzing the camera angle and pedestrian speed in the real world, it can be concluded that the displacement of center-of-mass coordinates meets the conditions:

2473-3547/16 $31.00 © 2016 IEEE
DOI 10.1109/ISCID.2016.29

$$-w_x \le \Delta X \le w_x \text{ , } -w_y \le \Delta Y \le w_y \tag{4}$$

$$w_x = W_{t-1} \text{ , } \quad w_y = H_{t-1} \text{ ,} \tag{5}$$

where $W_{t-1}$ and $H_{t-1}$ respectively are the width and height of the targets boundary in the frame $t$-1 , $w_x$ and $w_y$ are the maximum range of movement from the initial position. The closeness degree of the target centroid is defined as:

$$D(i,j) = \frac{\sqrt{(X_t - X_{t-1})^2 + (Y_t - Y_{t-1})^2}}{W_{t-1} * H_{t-1}} \tag{6}$$

where $D(i,j)$ is the closeness degree of the centroid position $(X_t,Y_t)$ and $(X_{t-1},Y_{t-1})$ . The smaller the $D(i,j)$ is, the larger possibility that target $i$ and $j$ are the same one .

### B. Area Matching

The video sequences are captured by a 25 Hz camera in which the interval of two image frames is 40 ms. Normally, the motion changes in 40ms are subtle and the size of the target area is invariable in two adjacent frames. Therefore, the closeness degree of area can also measure the similarity of the targets in two images. We use the method in [11] to get the area of the target and establish the model as

$$S(i,j) = \frac{|A_t - A_{t-1}|}{A_{t-1}} \tag{7}$$

where $S(i,j)$ is used to measure the similarity between the area $A_t$ and $A_{t-1}$ . The smaller the $S(i,j)$ is, the less difference of the area between target $i$ and target $j$ , the more similar they are.

### C. Corner matching

The corner of object refers to the point of which gray scale variation steeply in two-dimensional image or which is the maximum of curvature in image edge curve. We use the Moravec detector [12] to detect corners in the video sequences. The Moravec detector calculates the sum of squared differences(SSD) between a detected pixel and its adjacent pixels in the horizontal, vertical, and diagonal directions. We find that there are several corners extracted on the pedestrian, which may cause tracking errors and increase computation. Thus we process these points and find the suitable corners. All the position of corners are saved as $F(x_i,y_i)$ , and $P(x_j,y_j)$ is the position of selected suitable corners. We use a matching algorithm based on block [13] to achieve the motion estimation of the target corners. A $m*n$ image block is designed by taking the SSD between the image blocks and considered as a sliding template.In order to reduce the computation time, the sum of absolute differences(SAD) is taken between Blocks, which takes place of SSD, as calculated in

$$SAD(i,j) = \sum_{x=-m/2+1}^{m/2} \sum_{y=-n/2+1}^{n/2} | I_t(x,y) - I_{t-1}(x+i, y+j) |$$

$$i = -\frac{M}{2}+1,\ldots,\frac{M}{2}; j = -\frac{N}{2}+1,\ldots,\frac{N}{2}; \tag{8}$$

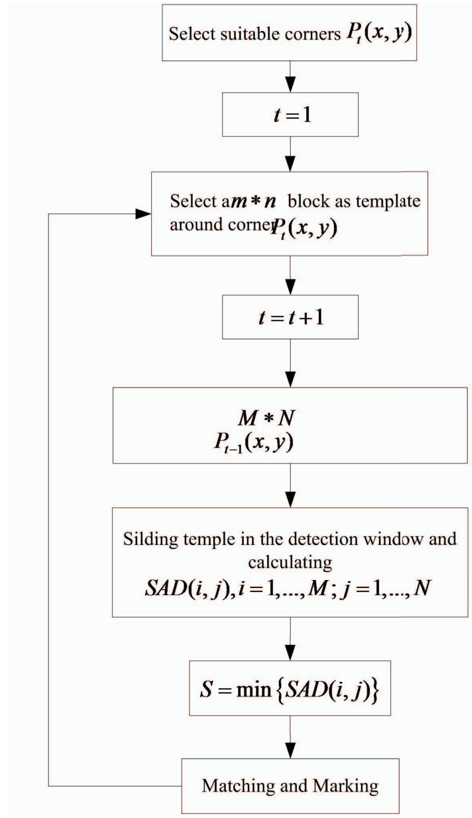


Figure 1. Flow chart of corner matching

where $I_t(x,y)$ is the pixel value of $(x,y)$ .The steps of corner matching are listed as Fig.1.

In this paper, the input values of the matching function are $D(i,j)$ , $S(i,j)$ and $SAD(i,j)$ with the corresponding weight coefficients $\alpha$ , $\beta$ and $\gamma$ . Therefore, the matching function can be defined as

$$V(i,j) = \alpha * D(i,j) + \beta * S(i,j) + \gamma * SAD(i,j) \tag{9}$$

where $D(i,j)$ , $S(i,j)$ and $SAD(i,j)$ are obtained from Eq.6, Eq.7 and Eq.8. The values of $\alpha$ , $\beta$ and $\gamma$ are decided by the different importance of each characteristic,which satisfied the equation of $\alpha + \beta + \gamma = 1$ . According to Eq.6 , Eq.7 and Eq.8, we find that when the values of $D(i,j)$ , $S(i,j)$ and $SAD(i,j)$ are minimum, the distance of target $i$ and $j$ is nearest and the two targets are most similar. Therefore, we conclude that when the value of $V(i,j)$ is minimum, target $i$ and $j$ are the best matching.

## III. PEDESTRIAN TRAJECTORY ANALYSIS

Pedestrian trajectory of the same target is obtained by processing continuous frames of video sequences. The position coordinates of pedestrian centroid are recorded when pedestrian target appears for the first time. The tracking sequences can be defined as

$$Track = \{(X_i,Y_i),(X_{i+1},Y_{i+1}),\ldots,(X_{i+n},Y_{i+n})\} \tag{10}$$

where $(X_i, Y_i)$ is the trajectory point in frame $i$ . $n$ is the number of trajectory points.

*A. Speed estimation*

The motion distance of pedestrian can be obtained by the start and end points of the trajectory in the real world , and we can get the motion time with matching times $n$ multiplied the interval $\Delta t$ of two adjacent frames. Then the speed expression is defined as

$$v = \frac{|s_n - s_0|}{n * \Delta t} \tag{11}$$

When $0.3 < v < 2$ , we consider this target as a pedestrian.

*B. Direction and position estimation*

The motion direction is an important information to detect pedestrian, which can confirm whether the pedestrian breaking into the road or walking along the road. Fig.2 shows the model of pedestrian motion direction and road direction.

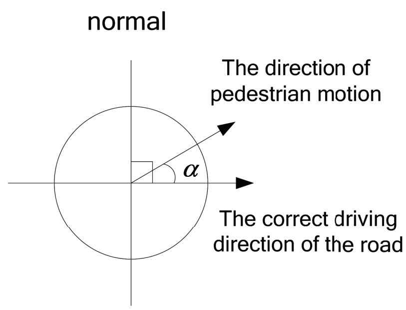



Figure 2. The graph of pedestrian motion direction.

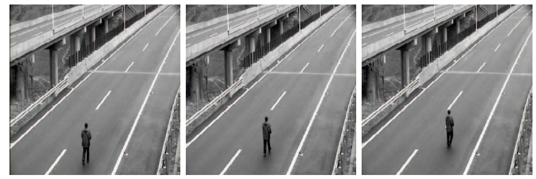
From the motion trajectory of pedestrian, we can obtain the motion vector $(R_x, R_y)$ . $(L_x, L_y)$ is the correct driving direction of the road.

$$\alpha = \left| ar\cos\left(\frac{R_x * L_x + R_y * L_y}{\sqrt{R_x^2 + R_y^2} + \sqrt{L_x^2 + L_y^2}}\right) \right| \tag{12}$$

$$C\_Direct = \begin{cases} true & \alpha > T \\ false & others \end{cases} \tag{13}$$

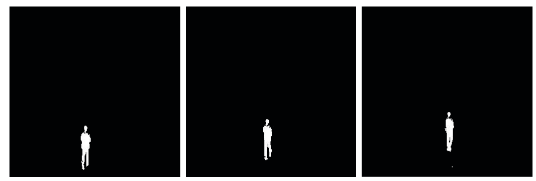
## IV. Experiment Results

We have tested our method in different complicated traffic scenarios to verify the validity. The experiments are conducted in Windows XP platform with a Intel Core i5-2450M(2.5GHz) processing unit and 4-GB random access memory. The size of each image is 720 × 280,and and the sampling frequency is 25 frames/s.The program implementation uses Visual C++ on a raw video format.



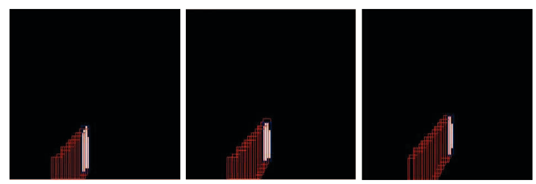
(a) (b) (c)

Figure 3. Gray-scale images of video sequences. (a), (b), (c) respectively represent the gray-scale images of 2000th frame, 2018th frame and 2038th frame.



(a) (b) (c)

Figure 4. Binary images of pedestrian targets. (a), (b), (c) respectively represent the binary images of 2000th frame, 2018th frame and 2038th frame.



(a) (b) (c)

Figure 5. Pedestrian targets matching and tracking. (a), (b), (c) respectively represent the tracking images of 2000th frame, 2018th frame and 2038th frame.

Fig.3, Fig.4 and Fig.5 show the tracking results of a highway scene in Chongqing province of China. The traffic flow density of this scene is small and external environment condition is relatively stable. Due to the images have good quality and less interference noise, pedestrian targets that we have obtained is pretty good shown as Fig.4. In Fig.5 where red marker boxes are targets in previous frames and blue marker box is in current frame, we can see that this method has a effective results of the continuous tracking of pedestrian targets.

(a) (b) (c)

Figure 6. Gray-scale images of video sequences. (a), (b), (c) respectively represent the gray-scale images of 5395th frame, 5396th frame and 5410th frame.

(a) (b) (c)

Figure 7. Binary images of pedestrian targets. (a), (b), (c) respectively represent the binary images of 5395th frame, 5396th frame and 5410th frame.

(a) (b) (c)

Figure 8. Pedestrian targets matching and tracking. (a), (b), (c) respectively represent the tracking images of 5395th frame, 5396th frame and 5410th frame.

Our method has also been applied in urban traffic video to verify the effectiveness.Fig.6, Fig.7 and Fig.8 show the tracking tests of a urban scene in Xi'an province of China. In Fig.8, (a) is the previous frame before pedestrians target loss in the process of matching track, (b) shows the failure of tracking because the target is occluded by vehicles, and (c) shows that pedestrian target appears on the 5410$^{th}$ frame again when vehicle passed and this target is regarded as a new target which can not achieve matching in 5410$^{th}$ frame.

COMPLEX TRAFFIC SCENES

| Scenes | Precision ratio | Omission ratio | Fall-out ratio |
|---|---|---|---|
| Highway | 97.93% | 1.8% | 0.27% |
| Urban Traffic | 96.02% | 3.4% | 0.58% |
| Tunnel | 94.89% | 4.7% | 0.41% |
| Night Time | 94.17% | 5.2% | 0.63% |

The final tests performance results are shown in Table Ⅰ. We can conclude that the average detection accuracy of our method is 95.8%. The positive detection rates of tunnel and night condition are both low, because of the influence of light. The proposed method demonstrated good performance on daytime, particularly on highways.

## V. CONCLUSION

In this paper, a method based on multi-feature is proposed for pedestrian tracking. At first, we construct a matching function with multi-feature according to the high similarity of the characteristics between two adjacent frames of a same target. Then pedestrian tracking are obtained by a matching mode which achieves the continuous tracking of targets. Last but not least, motion information ,such as speed motion trajectory. According to this method, we can give a warning classification for pedestrian events in traffic surveillance. The experimental results show that this method has good performance on tracking, which can overcome the impact of light sudden change and occlusions to a certain extent. In other complicated conditions including rain or nighttime,this method still has some limitation.We will continue to improve the accuracy of detection and strengthen the traffic analysis algorithms for the urban traffic surveillance in the future.


## ACKNOWLEDGMENT

The authors would like to thank the support of the Natural Science Foundation of Shaanxi Province (No. 2015JZ018, No.2015JQ6230).


## REFERENCES


[1] Junxian Wang, George Bebis, Ronald Miller, "Robust video-based surveillance by integrating target detection with tracking," IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshop, 2006.

[2] Fouad Bousetouanea, Lynda Diba, and Hichem Snoussib,"Robust detection and tracking pedestrian object for real time surveillance applications," International Conference on Graphic and Image Processing, 2011.

[3]Prioletti A., Mogelmose A.,Grisleri P., Trivedi M.M.,"Part-based pedestrian detection and feature-based tracking for driver assistance: real-time, robust algorithms, and evaluation," IEEE Trans. on Intelligent Transportation Systems, vol. 14, no. 3, pp. 1346-1359, Sept. 2013.

[4] Jifeng Ning, Wei Yu, Shuqin Yang,"An active contour tracking method by matching foreground and background simultaneously," IEEE Int. Conf. on Image Processing (ICIP),pp. 3944-3948, 2013.

[5] Ke-Chun Li, Hui-Chun Wang and Jing-Ming Chiu,"Pedestrian tracking system by using human shape prior model," IEEE Int. Conf. on Automation Science and Engineering, pp.18-22, Aug. 2014.

[6] Min Wang, Hong Qiao and Bo Zhang, "A new algorithm for robust pedestrian tracking based on manifold learning and feature selection," IEEE Trans. on Intelligent Transportation Systems, vol.12, no. 4, pp. 1195-1208, Dec. 2011.

[7] JongSeok Lim and WookHyun Kim, "Detecting and tracking of multiple pedestrians using motion, color information and the AdaBoost algorithm," Multimedia Tools and Applications, vol. 65, no. 1, pp. 161–179, July 2013.

[8] Tudor Barbu,"Pedestrian detection and tracking using temporal differencing and HOG features," Computers and Electrical Engineering, vol.40, no.4,pp.1072-1079, May 2014.

[9] Tung H. Dinh, Dao Q. Vu, Vu-Duc Ngo, Nam Pham Ngoc and Vu T. Truong,"High throughput FPGA architecture for corner detection in traffic images," IEEE Int. Cof. On Communications and Electronics (ICCE), pp. 297-302, 2014.

[10] Mohanna F. and Mokhtarian F., "A multi-scale approach to corner tracking," International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, vol.10, no. 3, pp. 74-81, Feb. 2002.

[11] Guo Wen-pu, Sun Ji-yin and Ren Jun, "Target match recognition method based on contour_area feature," IEEE Int. Conf. on Electrical and Control Engineering (ICECE), pp. 2225-2227, June 2010.

[12] H. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover," Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-80-03, 1980.

[13] Huan-Sheng Song, Sheng-Nan Lu, Xiang Ma, "Vehicle behavior analysis using target motion trajectories," IEEE Trans. on Vehicular