Optik 154 (2018) 22–32

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

Pedestrian abnormal event detection based on multi-feature fusion in traffic video

Xuan Wang*, Huansheng Song, Hua Cui

School of Information Engineering, Chang'an University, Xi'an, China

ARTICLE INFO

Article history:
Received 21 September 2016
Received in revised form 29 August 2017
Accepted 28 September 2017

Keywords:
Abnormal behavior event
Multi-feature fusion
Pedestrian trajectory analysis
Event classification

ABSTRACT

Pedestrian abnormal event detection is an active research area to improve traffic safety for intelligent transportation systems (ITS). This paper proposes an efficient method to automatically detect and track far-away pedestrians in traffic video to determine the abnormal behavior events. Firstly, pedestrian features are extracted by the multi-feature fusion method. Then, the similar features in current frame of all candidate objects are matched with the characteristic information of pedestrians in the previous frame which is considered as a template. Finally, pedestrian trajectory analysis algorithms are employed on the tracking trajectories and the motion information is attained, which can realize the early classification warning of pedestrian events. Experimental results on different traffic scenes in practice demonstrate that this method has good robustness in complex traffic. Moreover, the proposed method performs better compared with some other methods.

1. Introduction

In the past several years, the traffic abnormal events, such as illegal parking, abandoned object, illegal pedestrian, speeding, overloading, have been an increasing problem in traffic safety. Obviously, pedestrian abnormal behavior is one of the main reasons for the traffic accidents. In order to detect and manage illegal pedestrian events as soon as possible, this paper focuses on the issue of pedestrian abnormal behavior detection, which contains pedestrian detection, pedestrian tracking, and pedestrian trajectories analysis.

Pedestrian detection has attracted the attention of numerous researchers in recent years. It is an important issue in intelligent transportation systems (ITS). It requires preprocessing which consists of the recognition of the pedestrian target in video sequences and distinction of pedestrians from other video objects. The features of pedestrian is quite different from other objects, thus many researchers focus on the feature extraction to detect pedestrian. Gavrila and Giebel [1] presented a generic system for shape-based object recognition, which use the texture features to train a neural network to detect the pedestrian. Fang et al. [2] proposed a shape-independent pedestrian detection method in infrared images. It defines multi-dimensional histogram-, inertial-, and contrast-based classification features, which are shape-independent, complementary to one another, and capture the statistical similarities of image patches containing pedestrians with different poses. To avoid large areas of search, the system requires a pedestrian template and use brightness similarity to match candidate images and templates. Mohan et al. [3] presented a general example-based framework for detecting pedestrian by four components of the human body: the head, legs, left arm, and right arm. Xu et al. [4] adopted some local features of key parts of human

* Corresponding author.
E-mail address: jessica036@126.com (X. Wang).

https://doi.org/10.1016/j.ijleo.2017.09.104
0030-4026/

body to assist pedestrian detection. It used both appearance and motion global features of human body to select candidates, and then used local features of head and leg to do further confirmation. By combining Histograms of Oriented Gradients (HOG) and Local Binary Pattern (LBP) as the feature set, Wang et al. [5] proposed a pedestrian detection approach capable of handling partial occlusion, which use two kinds of detectors, i.e., global detector for whole scanning windows and part detectors for local regions, to learn from the training data by linear SVM.

Pedestrian tracking which locates the position of each detected target in the frames of analysed video sequences has a wide range of applications. In recent years, loads of methods for pedestrian tracking including active contour tracking [6], model tracking [7], feature tracking [8], etc., have been proposed by vision researchers. However, these approaches have some disadvantages. For example, active contour tracking is difficult to locate the contour of target accurately and mode tracking has a trouble in establishing pedestrian models which requires a large amount of computation. Because the objects of this paper are traffic video sequences, which are easily influenced by external environment, and traffic video incident detection system adopts mononuclear DSP processing chip which need to deal with a mass of image data in a short time. All of these require a high real-time ability, thus both of the first two methods are not suitable. However, features such as Gabor, color, edge, centroid, area and corner have a good performance in real-time tracking with less computation. Lim and Kim [9] tracked pedestrians with possibly partial occlusions with block matching method using color information. Barbu [10] used a robust pedestrian tracking method based on template matching process of Histogram of Oriented Gradient (HOG). Dinh et al. [11] presented a high throughput FPGA architecture for detecting corner features on traffic images.

Pedestrian trajectories analysis involves the analysis and recognition of motion patterns, and the production of high-level description of actions and interactions [12]. In general, it can be classified into two main categories of behavior description methods: statistical models and formalized reasoning. The Bayesian network model [13] is one of the statistical model. This model classifies certain events and behaviors by analysis of time sequences and statistical modeling. These methods need high-level reasoning based on a large amount of prior knowledge. Formalized reasoning [14] uses symbol systems to represent behavior patterns, and classified methods such as predication logic to recognize and classify events. For instance, Kojima et al. [15] proposed a method for generating natural language descriptions of pedestrian behaviors appearing in the image sequences.

In this paper, a method based on multi-feature fusion is used to detect and track pedestrians in traffic video so as to classify the pedestrian event. Motivated by that single feature can not solve the problem of segmentation very well, we combine several features of pedestrian and construct a mathematical model of matching function to detect pedestrian targets in traffic video. The proposed method has been tested on real-time video sequences and obtains good performance. The main contributions of this paper are concluded as follows:

1. We combine three different features of pedestrian to extract the pedestrian target. It avoids the limitations of a single feature, which has good robustness in complex traffic scene.
2. We establish the discriminant mode of pedestrian abnormal event to alarm these events timely in traffic scene.
3. The proposed method has better performance compared with some other method in the discrimination of pedestrian abnormal event.

The rest of the paper is organized as follows. Sections 2 and 3 describe the method of multi-feature extraction and fusion and the implementation of tracking, respectively. In Section 4, the motion information of pedestrian trajectory are analysed. Experiment results are given in Section 5. The final section draws the conclusion of this paper.

2. Multi-feature extraction and fusion

The pedestrian target of adjacent frames can be matched with certain characteristic information. In this paper, we bring some geometrical characteristics of pedestrian target together to establish the matching function as

$$F(x) = \sum_{i=1}^{n} w_i f_i(x), \tag{1}$$

where $f_i(x)$ is the entry data of $F(x)$, w_i is the corresponding weights, and $F(x)$ is defined as matching function. It is obvious that entries are the decisive parameters 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 entries including degree of location closeness, degree of area similarity and degree of corner point matching, and the weight coefficients are obtained by machine learning.

2.1. 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, so pedestrian obtained in video image is quite small. Assuming that the speed of pedestrians motion is uniform so that the movement distance of pedestrians is short during adjacent frames. What is more, the altitude

variation of pedestrians is approximately linear, thus centroid position is used to describe the similarity of pedestrian targets in video sequences. Then, the model of pedestrian motion is established as

$$\begin{aligned} M_t &= M_{t-1} + \Delta M \\ \Delta M &= H_t V_t + \xi_t, \end{aligned} \tag{2}$$

where $M_t = [X_t, Y_t]^T$ is the center of mass coordinate in tth frame, in which X_t is the horizontal axis and Y_t is the vertical axis. $\Delta M = [\Delta X, \Delta Y]^T$ is the position displacement of pedestrian target from $(t-1)$th frame to tth frame, in which ΔX is variation of horizontal displacement and ΔY is vertical displacement. $H_t = \Delta tE$ is the variation of time and $V_t = [V_x, V_y]^T$ is instantaneous speed of pedestrian targets, in which V_x is the projection speed in the horizontal direction and V_y is in the vertical direction. $\xi_t = [\xi_x, \xi_y]$ is the displacement error, in which ξ_x is in the horizontal direction and ξ_y is in the vertical direction.

According to (2), the displacement of targets can be divided into horizontal direction ΔX and vertical direction ΔY in the time interval Δt. Meanwhile, the movement of targets can be considered as a uniform motion in adjacent frames. In consideration of the errors, we can regard ΔX as the sum of the uniform motion displacement in Δt time and displacement error ξ_x, defined as $\Delta X = V_x \Delta t + \xi_x$. Likewise, ΔY is defined as $\Delta Y = V_y \Delta t + \xi_y$.

Analysing the camera angle and pedestrian speed in practice, it can be concluded that the displacement of center of mass coordinates should satisfy

$$\begin{aligned} &|\Delta X| \le w_x, \ |\Delta Y| \le w_y, \\ &w_x = W_{t-1}, \ w_y = H_{t-1}, \end{aligned} \tag{3}$$

where W_{t-1} and H_{t-1} are the width and height of the targets boundary in the $(t-1)$th frame and tth frame, respectively. w_x and w_y are the maximum range of movement from the initial position.

From the aforesaid threshold, it can be learned that if the difference value between centroid coordinates of the targets in the $(t-1)$th frame and the tth frame satisfies (4), the two targets are considered as the same target.

$$\begin{aligned} &|X_t - X_{t-1}| \le W_{t-1} \\ &|Y_t - Y_{t-1}| \le H_{t-1}. \end{aligned} \tag{4}$$

In this paper, the closeness degree of the target centroid is an entry data of the (1), thus it is necessary to establish a model to measure the closeness degree of the target centroid. It is defined as

$$D(i,j) = \frac{\sqrt{(X_t - X_{t-1})^2 + (Y_t - Y_{t-1})^2}}{W_{t-1} H_{t-1}}, \tag{5}$$

where $D(i,j)$ is the closeness degree of the centroid position (X_t, Y_t) of target i in the tth frame and the centroid position (X_{t-1}, Y_{t-1}) of target j in the $(t-1)$th frame. The smaller the $D(i,j)$ is, the larger possibility the two targets belong to the same one are.

2.2. Area

The video sequences are captured by a 25-Hz camera in which the interval of adjacent frames is 40 ms. Normally, the motion changes in 40 ms are subtle and the size of the target area is invariable in adjacent frames. The area calculation method is to count the number of pixels within the boundary. The area of the target in this paper is represented by the number of valid image blocks (binary image blocks) contained within its boundaries. Thus, the area of the pedestrian is

$$A = \sum_{x=left}^{right} \sum_{y=bottom}^{top} f(x,y), \tag{6}$$

where $f(x, y)$ is the image block, whose gray value is 255. Therefore, the closeness degree of area can also measure the similarity of the targets in two images, the model is built 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 of target i in the tth frame and the area A_{t-1} of target j in the $(t-1)$th frame. The smaller the $S(i,j)$ is, the less difference of the area between target i and target j and the more similar they are.

2.3. Feature corner

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 [16] tor detect corners in the video sequences. Moravec detector calculates the sum of squared differences (SSD) between a detected pixel and its adjacent pixels in the

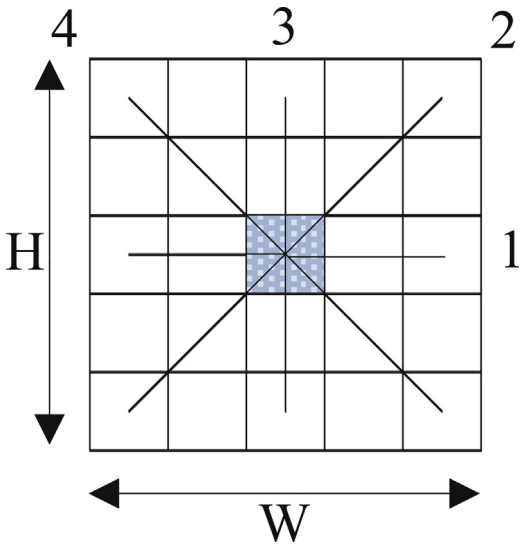


Fig. 1. Improved frame difference method.

horizontal, vertical, and diagonal directions. However, this paper uses the method based on block to achieve the target corners. A $m \times n$ 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. SAD measures the difference of the detected pixel (x, y) from surrounding image patches.

Four SAD values corresponding to four directions, i.e. number 1, 2, 3 and 4 in Fig. 1, are obtained by (8), and the smallest one is regarded as the $V(x, y)$. If $V(x, y)$ of the center point (x, y) is greater than a threshold value T, this point is marked as a corner. Here, $W \times H$ is the size of the image patch and $w = \text{int}(W/2)$, $h = int(H/2)$.

$$\begin{aligned} SAD_1 &= \sum_{a=-w}^{w}\sum_{b=-h}^{h} |g(x-i+a, y+b) - g(x+i+a, y+b)| \\ SAD_2 &= \sum_{a=-w}^{w}\sum_{b=-h}^{h} |g(x-i+a, y-j+b) - g(x+i+a, y+j+b)| \\ SAD_3 &= \sum_{a=-w}^{w}\sum_{b=-h}^{h} |g(x+a, y-j+b) - g(x+a, y+j+b)| \\ SAD_4 &= \sum_{a=-w}^{w}\sum_{b=-h}^{h} |g(x+i+a, y-j+b) - g(x-i+a, y+j+b)| \end{aligned} \tag{8}$$

$$V(x, y) = \min\{SAD_1, SAD_2, SAD_3, SAD_4\}. \tag{9}$$

where $g(x, y)$ is the gray value of point (x, y). It is can be found that there are several corners extracted on the pedestrian, which may cause tracking errors and increase computation. Thus we need process these points and find the suitable corners. The specific steps of corner selecting are listed as Fig. 2. All the position of corners are saved as $I(x_i, y_i)$, where $i \in \{1, 2, \times, N\}$ and N is the total account of corners, $P(x_j, y_j)$ is the position of selected suitable corners. L is the distance between $I(x_i, y_i)$ and $P(x_j, y_j)$. If L is larger than a threshold η, $I(x_i, y_i)$ is considered as a suitable corner and is save in $P(x_j, y_j)$. In addition, η is the height of pedestrian. Consequently, we can get suitable corners of each pedestrian in the tth frame.

The specific matching steps are listed as Fig. 3. $P_t(x, y)$ is the corner position in the tth frame. It is used as the center of sliding template. In the $(t+1)$th frame, we select a detection window around the corner of previous frame and calculate the $C(i, j)$ in each sliding block.

$$C(i, j) = \sum_{a=1}^{w}\sum_{b=1}^{h} \frac{|P_t^i(a+x, b+y) - P_{t+1}^j(a+x, b+y)|}{P_t^i(a+x_t, b+y_t)}, \tag{10}$$

where $C(i, j)$ is the similarity degree of corner i in the tth frame and corner j in the $(t+1)$th frame. The smaller the $C(i, j)$ is, the more similar the two corners are. Finally, the corners in the tth frame are matched to the $(t+1)$th frame.

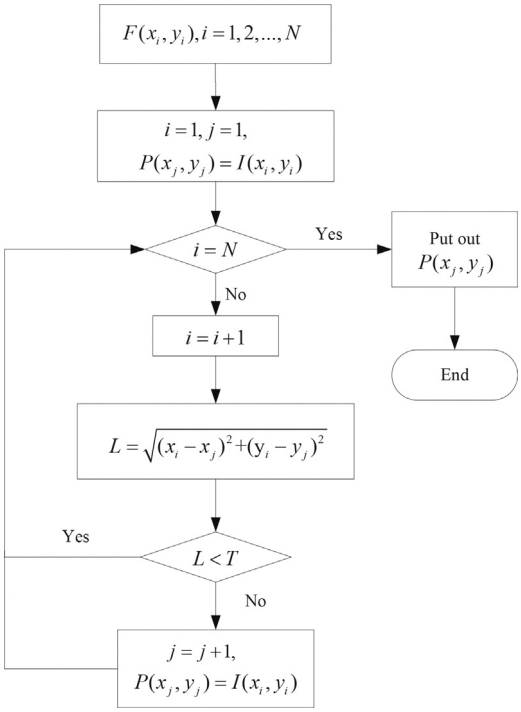


Fig. 2. Selection of the appropriate corners.

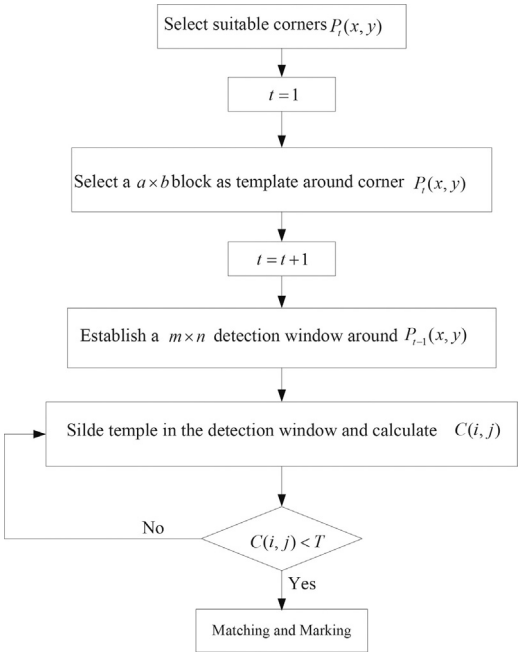


Fig. 3. Procedure of matching the corners.

3. Pedestrian tracking

In this paper, the entries of function are $D(i, j)$, $S(i, j)$ and $C(i, j)$ with the corresponding weight coefficients α, β and γ. Therefore the matching function is defined as

$$F(i, j) = \alpha D(i, j) + \beta S(i, j) + \gamma C(i, j), \tag{11}$$

(a) (b) (c)

Fig. 4. Tracking trajectories of pedestrian. (a) Highway scene, (b) urban road scene, (c) tunnel scene.

where $D(i, j)$, $S(i, j)$ and $C(i, j)$ are obtained from Eq. (5), (7) and (10), respectively. α, β and γ are decided by the different importance of each feature, which satisfied the equation of $\alpha + \beta + \gamma = 1$. For each target i in the tth frame, it can be found a target j in the $(t+1)$th frame, which makes $F(i, j)$ minimum. Then, it is converted to a optimization problem to find the minimum of $F(i, j)$:

$$\min F(i, j) \text{ subject to } j \in \{1, 2, \ldots, M\}, \tag{12}$$

where M is the number of the pedestrian target in the $(t+1)$th frame. Therefore, it is concluded that if the value of $F(i, j)$ is smallest, target i and j are the best matching.

Fig. 4 shows that the proposed method is employed in different traffic videos to verify the effectiveness. The matching method, which is based on multi-feature fusion, can realize the pedestrian tracking in many different scenes, and it also can effectively solve the problems of object segmentation, such as target is imperfect or missing.

4. 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 enters into the traffic scene for the first time. The trajectory points can be defined as

$$Track = \{(x_i, y_i), (x_{i+1}, y_{i+1}), \ldots, (x_{i+n}, y_{i+n})\}, \tag{13}$$

where (x_i, y_i) means that the target first entry into the traffic scene is in the ith frame and its position is record as (x_i, y_i). n is the number of trajectory point that the same target is matched continuously in video sequence.

4.1. Trajectory pretreatment

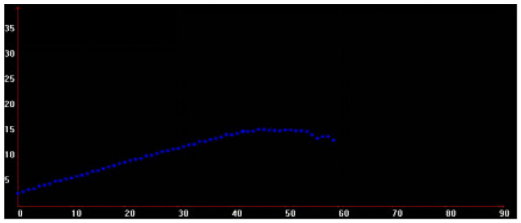
Trajectory pretreatment mainly contains the trajectory screening and the trajectory points mapping.

We screen the trajectories that have been obtained from video sequences and leave some trajectories which contain sufficient motion information. Firstly, T_1 is set for every trajectory, which is a threshold for n. If $n > T_1$, this trajectory is considered to have enough motion information and is reserved for further management. After that, the horizontal and vertical movement distances of every trajectory are calculated from beginning to end in the image, which are defined as L_x and L_y, respectively. T_2 and T_3 are set as thresholds for L_x and L_y. If $L_x > T_2$ & $L_y > T_3$, this trajectory is considered as effective trajectory. Trajectory pretreatment can remove some interference trajectories. In this paper, the value of T_1, T_2, and T_3 is respectively 20, 40, 30.

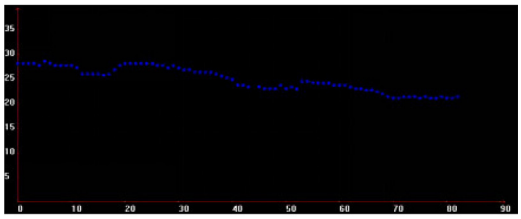
Because camera imaging is not a linear model, the trajectory on image can not reflect the actual motion information of pedestrian due to image distortion which results from the translation of the world coordinate system (WCS) and image coordinate system (ICS). However, the mapping table can reflected the corresponding relationship from image pixels to the actual distance. Therefore, we can convert the trajectory points to the actual distance of target movement, as shows in

$$Track' = \{(S_t, t), (S_{t+1}, t+1), \ldots, (S_{t+n}, t+n-1)\}, \tag{14}$$

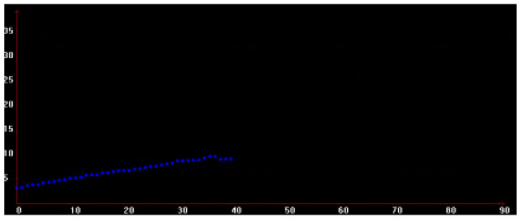
where (S_t, t) is the corresponding relationship between actual distance and time frames, which means that the target first enter into the traffic scene is in the tth frame and the actual distance is S_t. $Track'$ is the actual movement of pedestrian target, which consists of n groups relationship of target actual movement distance and time frames. Fig. 5 shows the curve graphs of actual trajectory, which justified the method we proposed can track the pedestrian objects effectively, and the smooth movement curve further explain the actual motion state of pedestrians.



(a)



(b)



(c)

Fig. 5. Coordinate transformation of tracking trajectories. (a) Highway scene, (b) urban road scene, (c) tunnel scene.

4.2. Extraction of motion information

Through the above process, we reserve the trajectories which can reflect the motion feature of the object exactly. Then, further analysis of these trajectories can realize the detection of pedestrian abnormal events.

4.2.1. 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 of adjacent frames. Then the speed expression is defined as

$$v = \frac{|S_{t+n} - S_t|}{n\Delta t} \tag{15}$$

If $\zeta < v < \eta$, it is considered the target as a pedestrian. Here, ζ and η are the minimum and maximum velocity of pedestrian respectively, i.e., $\zeta = 0.3$ m/s, $\eta = 2.0$ m/s.

4.2.2. Direction and position estimation

The motion direction is an important information to detect pedestrian abnormal behavior, which can confirm whether the pedestrian breaking into the road or walking along the road. Fig. 6 shows the model of pedestrian motion direction and road direction.

From the motion trajectory of pedestrian, the motion vector V_p can be obtained by the linear fit of trajectory points. R is the correct driving direction of the road.

$$\theta = \left| \arccos \frac{V_p^T R}{\|V_p\|_2 + \|R\|_2} \right|. \tag{16}$$

$$P_{direction} = \begin{cases} True & \theta \le \phi \\ False & otherwise \end{cases} \tag{17}$$

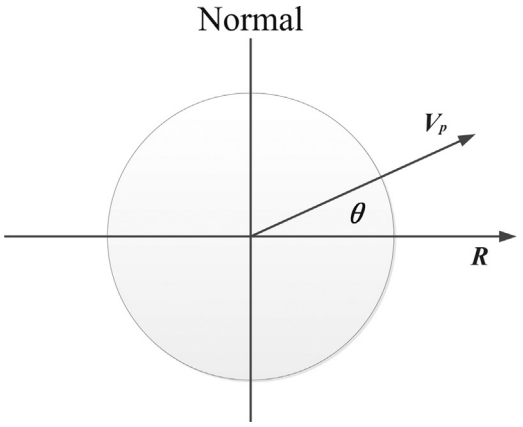


Fig. 6. Pedestrian motion direction.

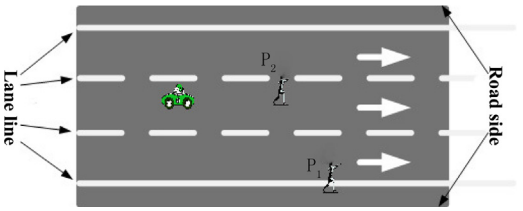


Fig. 7. Pedestrian motion direction.

As shown in (17), if $\theta \leq \phi$, it is considered that pedestrian breaking in the road which seriously threats the normally running of vehicles. Then, the alarm is triggered. In practice, the motion direction of pedestrian is random, and this paper use the threshold $\phi = 30^{\circ}$.

4.2.3. Pedestrian events classification

The motion information of pedestrians is used for realizing early warning and confirming to evaluate the risk of pedestrian events. As Fig. 7 shows, the pedestrian P_1 is out of the driveway where is the temporary realm of vehicles. It normally does not allow other objects expect vehicles to enter. The pedestrian P_2 is on the inside lane, which seriously influences the normally running of vehicles, whose risk degree is higher than P_1. Then, the position model is defined as

$$P_{position} = \begin{cases} True & \text{Out of the driveway} \\ False & \text{Within the inside lane} \end{cases} \tag{18}$$

In this paper, the early risk warning of pedestrian events are classified into three warning level: The high level is the most dangerous, i.e., P_2 in Fig. 7, which may lead to a fatal traffic accident; The mid level is middle dangerous, in which the pedestrian is moving toward the lane, and this situation will be developed into the high level in the future, i.e. P_1 in Fig. 7; The low level is the least dangerous, in which the pedestrian is walking outside the lane and the moving direction is not toward the lane. The classification of the pedestrian events are defined as

$$Level = \begin{cases} High & P_{position} = False \\ Mid & P_{position} = True \& P_{Direction} = False \\ Low & Others \end{cases} \tag{19}$$

5. Experiments

The propose method has been tested in different complicated traffic scenes to verify the robustness and effectively. The experiments are conducted in Windows 10 platform with a Intel Core i5-2450M (2.5 GHz) processing unit and 8-GB random access memory. The image has a size of 720 × 280, and its sampling frequency is 25 frames/s. The program implementation uses Visual C++ on a raw video format.

The experimental results of different traffic scenes, such as highway, urban road, tunnel, and night road, are performed in Fig. 8. The event information of Fig. 8(a) is that image coordinate of the pedestrian is (128,600), i/e., $P_{position} = False$, the direction of the pedestrian is from the road center to the side of the road, i.e., $P_{direction} = False$. The warming level of the pedestrian target in Fig. 8(a) is high. Likewise, the event information of other scenes can be obtained as shown in Table 1.

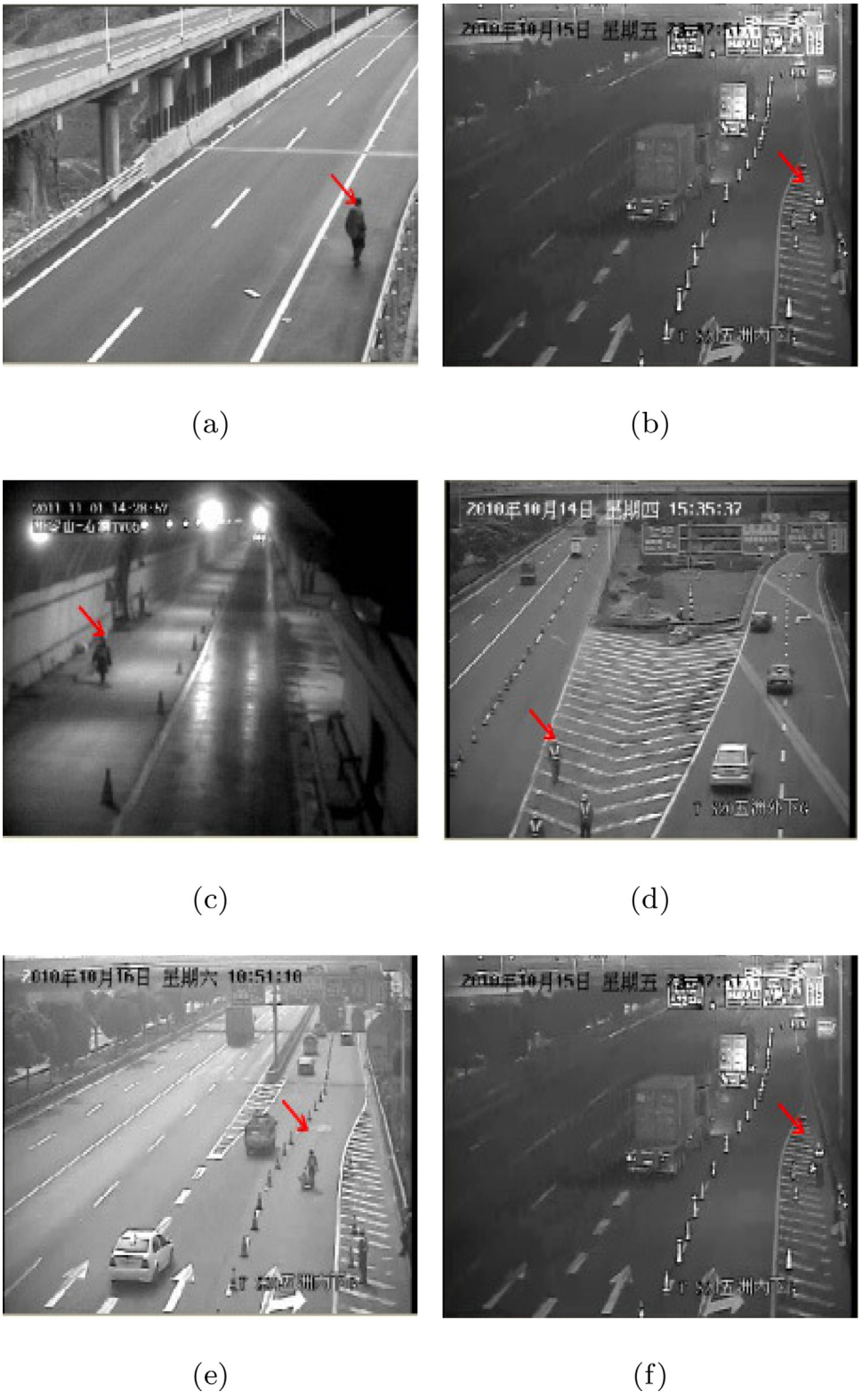


Fig. 8. Experimental results of different traffic scenes.

Table 1
Experiment results in different scenes.

Scenes	(x, y)	$P_{position}$	$P_{direction}$	*Level*
Chongqing Highway	(128,600)	*False*	*False*	*High*
Xi'an South Second Road	(228,536)	*True*	*True*	*Low*
Huayingshan Tunnel	(184,176)	*False*	*True*	*High*
Shanghai Highway	(80,192)	*False*	*False*	*High*
Shanghai Highway	(144,528)	*True*	*False*	*Mid*
Night of Shanghai Highway	(168,624)	*True*	*True*	*Low*

Table 2
Detection errors.

Scenes	FPR[a]	MR[b]	PR[c]
Chongqing Highway	0.27%	1.8%	97.93%
Xi'an South Second Road	0.58%	3.4%	96.02%
Huayingshan Tunnel	0.41%	4.7%	94.89%
Shanghai Highway	0.22%	2.3%	97.48%
Shanghai Highway	0.15%	1.3%	98.55%
Night of Shanghai Highway	0.63%	5.2%	94.17%

[a] FPR is the false positive rate.
[b] MR is the missing rate.
[c] PR is the precision rate.

Table 3
Detection results of Shanghai Hokkaido viaduct in practice.

Camera number	PN[a]	AN[b]	FA[c]	PR
01	4	4	0	100%
02	35	36	1	97.14%
03	2	2	0	100%
04	1	1	0	100%
05	2	2	0	100%
06	43	45	2	95.34%

[a] PN is the total number of the pedestrian.
[b] AN is the total number of alarm.
[c] FA is the number of false alarm.

Table 4
Comparison with other method.

Methods	FPR	MR	PR
[17]	6.95%	7.31%	85.74%
[18]	3.89%	6.75%	89.36%
Our method	0.58%	3.4%	96.02%

As shown in Tables 1 and 2, it can be seen that the proposed method of pedestrian abnormal detection in this paper is available for the illegal pedestrian events in various traffic scenes, such as highway, urban road and tunnel. Even in the night, the method still can accurately detect pedestrian events as long as the light illumination is enough. More specifically, the average detection accuracy of our method can reach 96%. Both of tunnel and night condition have low positive detection rates, that is because the light condition is poor. But it has good performance on daytime, particularly on highways. The above experimental results show that the proposed method has a strong robustness in complex traffic scenes. In addition, all these abnormal events can be detected within 5 s.

Moreover, in order to test the reliability of the proposed method in practice, it has been applied in the scene of Shanghai Hokkaido Viaduct for the whole day test. The test is divided into six sections and the test period are 45 days. The complexity of the traffic environment and the probability of the pedestrian are different in different sections. From Table 3, it can be seen that the average accuracy of pedestrian abnormal detection is 96.55%, which can effectively realize the illegal pedestrian warning in traffic scene. Meanwhile, the proposed method can work normally in the environment of night, cloudy, rain and so on.

Furthermore, the proposed method has been compared with other methods as given in Table 4. These methods are applied in the traffic video of the Xi'an South Second Road. In addition, Zhang and Liu [17] used the distance between contour points and the centroid to build the SVM classifier to analyse the behavior of pedestrian. Jiang et al. [18] built a behavior model between the pedestrian trajectory and the road to classify the different kinds of pedestrian abnormal behaviors. However, both of these methods have some weakness in the stage of pedestrian detection. In this paper, we combine the idea of the

former two papers and use the feature fusion to obtain the pedestrian target, which has good robustness in complex traffic scenes. It can be seen that our method significantly outperforms the two methods.

6. Conclusion

In this paper, a method based on multi-feature fusion is proposed to detect and track pedestrian for the illegal pedestrian event detection. At first, we construct a matching function with multi-feature based on the high similarity of the characteristics between adjacent frames. Then pedestrian tracking trajectories are obtained by a matching mode which achieves the continuous tracking of targets. Last but not least, motion information, such as speed and pedestrian movement direction, are gained by analysing motion trajectory. Based on this method, we can give a warning classification model for pedestrian abnormal events in traffic video. The experimental results show that the proposed method has better performance compared with other two methods, which can overcome the sudden change impact of light and occlusions to a certain extent. Moreover, the proposed method has a good robustness in different traffic scenes in practice. In other complicated conditions including rain or night time, the method still has some limitations. We will continue to improve the accuracy of detection and strengthen the traffic analysis algorithms for the urban traffic surveillance in the future.

Acknowledgements

This work is supported by the National Natural Science Fund of China (No. 61572083), the Natural Science Foundation of Shaanxi Province (No. 2015JZ018).

References

[1] D.M. Gavrila, J. Giebel, Shape-based pedestrian detection and tracking, in: Intelligent Vehicle Symposium. 2002. IEEE, vol. 1, 2002, pp. 8–14.
[2] Y. Fang, K. Yamada, Y. Ninomiya, B.K.P. Horn, I. Masaki, A shape-independent method for pedestrian detection with far-infrared images, IEEE Trans. Veh. Technol. 53 (2004) 1679–1697.
[3] A. Mohan, C. Papageorgiou, T.A. Poggio, Example-based object detection in images by components, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 349–361.
[4] Y.W. Xu, X.B. Cao, H. Qiao, Pedestrian detection with local feature assistant, in: 2007 IEEE International Conference on Control and Automation, 2007, pp. 1542–1547.
[5] X. Wang, T.X. Han, S. Yan, An HOG-LBP human detector with partial occlusion handling, in: 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 32–39.
[6] J. Ning, W. Yu, S. Yang, An active contour tracking method by matching foreground and background simultaneously, in: 2013 IEEE International Conference on Image Processing, 2013, pp. 3944–3948.
[7] K.-C. Li, H.-C. Wang, J.-M. Chiu, Pedestrian tracking system by using human shape prior model, in: 2014 IEEE International Conference on Automation Science and Engineering (CASE), 2014, pp. 1139–1143.
[8] F. Mohanna, F. Mokhtarian, A multi-scale approach to corner tracking, in: WSCG, 2002.
[9] J.S. Lim, W.H. Kim, Detecting and tracking of multiple pedestrians using motion, color information and the adaboost algorithm, Multimedia Tools Appl. 65 (2012) 161–179.
[10] T. Barbu, Pedestrian detection and tracking using temporal differencing and HOG features, Comput. Electr. Eng. 40 (2014) 1072–1079.
[11] T.H. Dinh, D.Q. Vu, V.-D. Ngo, N.P. Ngoc, V.T. Truong, High throughput fpga architecture for corner detection in traffic images, in: 2014 IEEE Fifth International Conference on Communications and Electronics (ICCE), 2014, pp. 297–302.
[12] W. Hu, T. Tan, L. Wang, S.J. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Trans. Syst. Man Cybern. C 34 (2004) 334–352.
[13] P. Remagnino, T. Tan, K.D. Baker, Agent orientated annotation in model based visual surveillance, in: ICCV, 1998.
[14] M. Mohnhaupt, B. Neumann, On the use of motion concepts for top-down control in traffic scenes, in: ECCV, 1990.
[15] A. Kojima, M. Izumi, T. Tamura, K. Fukunaga, Generating natural language description of human behavior from video images, in: ICPR, 2000.
[16] H.P. Moravec, Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover, Stanford University, 1980.
[17] J. Zhang, Z. Liu, Detecting abnormal motion of pedestrian in video, in: 2008 International Conference on Information and Automation, 2008, pp. 81–85.
[18] Q. Jiang, G. Li, J. Yu, X. Li, A model based method of pedestrian abnormal behavior detection in traffic scene, in: ISC2, 2015.