

Research Article

Dempster-Shafer Multifeature Fusion for Pedestrian Detection

**Hua Cui,¹ Lingling Peng,¹ Huansheng Song,¹ Guofeng Wang,²
Jiancheng Li,² Lu Guo,¹ and Chao Yuan¹**

¹ School of Information Engineering, Chang'an University, Xi'an 710064, China

² China Highway Engineering Consulting Corporation, Beijing 100097, China

Correspondence should be addressed to Hua Cui; huacui1122@gmail.com

Received 25 August 2014; Revised 4 November 2014; Accepted 5 November 2014

Academic Editor: Hongwei Guo

Copyright © Hua Cui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pedestrian detection is of great importance for ensuring traffic safety. In recent years, many works employing image-based shape features to recognize pedestrians have been reported. However, previous pedestrian detectors were in many cases not sufficient to achieve satisfactory results under complex weather conditions and complex scenarios. As a solution this paper exploits two video-based motion feature descriptors and applies such motion features to the detection task in addition to four classical shape features with the aim of significantly improving the detection performance. Our motion features are defined as the trajectory smoothness degree and motion vector field, which are derived from our proposed point tracking strategy beyond tough target segmentation. And then the appealing Dempster-Shafer theory of evidence (D-S theory) is applied to fuse these features, due to the fact that D-S theory is better than the classical Bayesian approach in handling the information with lack of prior probabilities. The proposed automatic pedestrian detection algorithm is evaluated on real data and in real traffic scenes under various weather conditions. Theoretical analysis and experiment results consistently show that the proposed method outperforms SVM-based multifeature fusion approach for pedestrian detection in terms of recognition ability and robustness in various real traffic scenes.

1. Introduction

The ability to detect and recognize human being is fundamental for minimizing the number of deaths in car accidents. In recent years the pedestrian detection has attracted more and more attention of researchers all over the world. One of the key problems in pedestrian recognition is how to extract robust and effective features from traffic surveillance videos. Many pedestrian detection algorithms have been put forward based on various features [1–8], such as shape features [1–3] and motion features [4, 5]. Tracking techniques are often used in pedestrian detection as well. In [8], pedestrian objects were tracked, and their shapes were recursively estimated to detect them.

However, previous pedestrian detectors are in many cases not sufficient to achieve satisfactory results. One essential cause is that the variability of the appearances and poses of the pedestrians makes the shape features unstable. Meanwhile, most of previous works are based on the accurate image

segmentation. Some of them employed image segmentation technologies to extract pedestrian hypotheses from images to calculate their shape features. Some executed image segmentation technologies to extract pedestrian targets from video frames and then tracked the segmented-pedestrian-region to obtain the motion features, whereas it is absolutely a tough task to achieve satisfactory image segmentation. Moreover, various complex weather conditions [9] the varying illumination, shadow effect, interference of other objects, and the incredibly unpredictable background of a traffic scene always make the pedestrian segmentation more challenging [10], leading to the unreliability of segmentation-based shape and motion features. As a solution this paper exploits two video-based motion feature descriptors and then applies such motion features to the pedestrian detection in addition to four classical shape features with the aim of significantly improving the detection performance. Our motion feature descriptors are defined as the trajectory smoothness degree and the motion vector field, which were derived from our

proposed point tracking strategy without target segmentation procedure.

Many research findings available show that using individual features to identify pedestrian targets tends to give an uncertain or even wrong result. On the other hand, different features contain different discriminative information and they are complementary to each other. If all these different features are utilized together and fused efficiently, the recognition performance is bound to be improved greatly. Multifeature fusion can produce a better global understanding of the observed pedestrian by decreasing the uncertainty related to the individual features, thus resulting in a well-adaptive answer to pedestrian detection. In fact, information fusion has already been widely applied in diverse fields [11–15] and there have been growing interests in its use in intelligent transportation systems [13–15] and in pedestrian detection [3, 6, 13, 14]. In [14], the two types of data, that is, the infrared video data and the lidar data, were considered, and invariant vectors were devised to distinguish pedestrians from other targets using support vector machines (SVMs). But SVM-based fusion approaches always involve great computational complexity due to their learning on large training data sets. To circumvent the problems mentioned above, this paper used the Dempster-Shafer theory of evidence (D-S theory) [16] to perform the multifeature fusion to establish a pedestrian recognition method, which could realize the pedestrian detection with high certainty and precision.

As the generalization of Bayesian theory, D-S theory is a powerful and efficient approach for solving data fusion problems in practical applications [11, 12, 15]. Moreover, unlike the Bayesian theory and SVM fusion, D-S theory has significant advantages in terms of the representation, measure, and fusion of the imprecise and uncertain information [16]. Therefore, this paper selects D-S theory to fuse the diverse features to increase the reliability of the pedestrian recognition. Additionally, besides four popular shape features of pedestrians, two efficient motion features are exploited to improve the detection robustness to the diversity of illumination, traffic scenes, pedestrian appearance, and so on.

In this paper, our major goal is to develop a pedestrian detection method based on multifeature fusion, which could provide a better robustness and higher detection correctness rate and lower false alert rate than pedestrian detection methods available. Our proposed method is derived by the aid of D-S theory along with our defined motion features. The rest of this paper is organized as follows. Section 2 reviews the D-S theory. And then Section 3 describes diverse features used to characterize and separate the objects of interest (e.g., the vehicles and pedestrians). Our proposed motion features especially are defined and developed concretely. In Section 4 our proposed multifeature fusion pedestrian detection method is presented step by step, and its performance is evaluated in real traffic scenes in Section 5, and experiment results verify the superiority of our proposed pedestrian detection method. Finally, Section 6 is dedicated to our conclusions.

2. D-S Theory

As each feature brings a piece of evidence on the presence of a pedestrian, the most of features do not discriminate pedestrians alone. To benefit from all information involved in diverse features, we need to combine them using a certain data fusion approach. D-S theory [16] is the generalization of the Bayesian theory and has been implemented to solve practical data fusion problems [11, 12, 15, 16]. Bayesian framework for data fusion makes an extensive use of prior probabilities, but the prior probabilities of some data are often unknown. D-S theory abandons the additive property of probabilities in classical probability frame and allows reasoning with weaker prior knowledge which is much easier to get. Moreover, unlike the classical Bayesian theory, it can handle the uncertain information and capture the imprecise nature of information. Consequently, the D-S theory appears as the best way to manage the imprecision of pedestrian features, reducing the uncertainty of the pedestrian detection.

In the D-S framework, evidence is assigned to the elements of the set. The set consists of all possible propositions and is called the discernment frame and often denoted by Θ , while $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$. A proposition about variable θ is any subset of Θ . 2^Θ is called the power set of Θ , denoting the set of all 2^n possible subsets of Θ . To give an example, let us think of a patient that can be affected by cancer or not. We can model this scenario by defining a variable C with the frame $\Theta = \{ac, nac\}$, where ac is the proposition “patient is affected by cancer,” nac the proposition “patient is not affected by cancer,” and $(ac \cup nac)$ the doubtful proposition “patient is or is not affected by cancer.” Each proposition is mapped onto a single subset and is assigned a basic belief mass through a basic probability assignment (BPA) function. The BPA function m over the propositions from 2^Θ has the following properties:

$$\begin{aligned} m : 2^\Theta &\longrightarrow [0, 1], & m(\emptyset) &= 0, \\ \sum_{A \in 2^\Theta} m(A) &= 1, \end{aligned} \quad (1)$$

where \emptyset denotes the empty set and the summation is taken over every possible subset A of Θ . The BPA function m maps the any subset A to a number $m(A)$, whose value is between 0 and 1. Each set A such that $m(A) > 0$ is called a focal element for 2^Θ . $m(A)$ is the part of probability that supports exactly A but does not support any strict subset of A due to lack of knowledge. In the previous example, if we had assigned mass 0.85 to $\{ac \cup nac\}$ proposition and 0.15 to $\{ac\}$, it would have meant that there is some evidence for the patient being affected by cancer, but based on current knowledge, a great part of our confidence cannot be assigned to none of the two specific propositions.

If we want to obtain the total belief for a set A , we must add the mass of all proper subsets of A plus the mass of A itself, thus obtaining the belief for the proposition A . The belief function is defined as

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B). \quad (2)$$

The belief function $\text{Bel}(A)$ contains all evidences attached to subsets of the proposition A and summarizes all our reasons to believe in A with the available knowledge. In D-S theory, information about each proposition is represented by an interval, bounded by two values, namely, the belief and the plausibility. The plausibility function is defined as follows:

$$\text{Pl}(A) = 1 - \text{Bel}(\bar{A}) = \sum_{B \cap A \neq \Phi} m(B). \quad (3)$$

The plausibility function $\text{Pl}(A)$ is the sum of all the masses that intersect the set of interest A . From the definition of $\text{Bel}(A)$ in formula (2), $\text{Bel}(\bar{A})$ implies correspondingly that we believe A is false. Therefore, $\text{Pl}(A)$ denotes the belief extent to which we believe A is not false. The interval $\text{Pl}(A) - \text{Bel}(A)$ represents the ignorance degree of A , that is, the lack of information (or the amount of doubt) about A . $\text{Bel}(A)$ and $\text{Pl}(A)$ are the lower-limit and upper-limit of belief degree of A , respectively, which usually be denoted by $A(\text{Bel}(A), \text{Pl}(A))$. For instance, for $A(0.25, 0.85)$, $\text{Bel}(A)$ and $\text{Pl}(A)$ are 0.25 and 0.85, respectively. Consequently, we believe that A is true with BPA mass of 0.25, and meanwhile we believe that A is false. By subtracting 0.85 from 1 we obtain the unbelief mass of A equal to 0.15. Thus we believe that A is true in a little bit more degree than that A is false.

In the case of imperfect data, fusion is a powerful approach to obtain more relevant information. After defining focal elements for each feature and assigning pieces of evidence to each focal set, the D-S theory provides an orthogonal fusion rule [16] to perform information fusion, that is, combining two BPA functions m_1 and m_2 to yield a new BPA function m . The BPA mass of the proposition C , resulting from the combination of two sources 1 and 2, is expressed as follows:

$$m(C) = m_1 \oplus m_2(C) = \begin{cases} 0, & C = \emptyset \\ \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{1 - K}, & C \neq \emptyset, \end{cases} \quad (4)$$

where K is defined as follows:

$$K = \sum_{A \cap B = \Phi} m_1(A) m_2(B). \quad (5)$$

K measures the conflict degrees between m_1 and m_2 . The bigger K corresponds to the higher conflict. If $K = 1$, it implies that the propositions m_1 and m_2 are incompatible, whereas the fusion of m_1 and m_2 will produce a new BPA function m when $K \neq 1$. Furthermore, Dempster's combination rule treats conflict as a normalization factor, so its presence is no longer visible after fusion.

Continuing the previous example, suppose that two doctors could provide information after their examining the patient and lead us to write the following BPA values:

$$\begin{aligned} m_1(\theta) &= \begin{cases} 0.8 & \theta = \{\text{ac}\} \\ 0.2 & \theta = \{\text{nac}\}, \end{cases} \\ m_2(\theta) &= \begin{cases} 0.3 & \theta = \{\text{ac}\} \\ 0.7 & \theta = \{\text{ac} \cup \text{nac}\}. \end{cases} \end{aligned} \quad (6)$$

Note that the first doctor is a cancer specialist and the second is not. The information the second doctor provides is quite limited about cancer diagnosis with mass 0.3, while most of the mass (0.7) is assigned to doubt. Fusing the two pieces of information according to the rule as presented in (5) results in

$$m(\theta) = \begin{cases} \frac{0.8 * 0.3 + 0.8 * 0.7}{1 - 0.2 * 0.3} = 0.8511 & \theta = \{\text{ac}\} \\ \frac{0.2 * 0.7}{1 - 0.2 * 0.3} = 0.1489 & \theta = \{\text{nac}\}. \end{cases} \quad (7)$$

It is seen that values after fusion are not far from those already assigned by m_1 , because the second doctor did not bring a clear contribution to the cancer diagnosis. Meanwhile, because the first doctor strong believes that this patient is affected by cancer (0.8) and assigns mass 0.2 to show her/his low belief of absence of cancer, little conflict is observed as $K = 0.2 * 0.3 = 0.06$.

Dempster's combination rule can be generalized to more than two hypotheses, and the belief function resulting from the combination of J information sources S_j is defined as

$$\begin{aligned} m(C) &= m_1(S_1) \oplus m_2(S_2) \oplus \dots \oplus m_J(S_J) \\ &= \frac{\sum_{\cap_{j=1}^J S_j = C} \left(\prod_{i=1}^J m_i(S_j) \right)}{1 - \sum_{\cap_{j=1}^J S_j = \emptyset} \left(\prod_{i=1}^J m_i(S_j) \right)}, \end{aligned} \quad (8)$$

where S_1, S_2, \dots, S_J are focal elements. However, it yields quite large computation burden. Therefore, we adopted an equivalent method to calculate multievidence fusion result. That is, we fused every two evidences using the combination rule described as formula (4) so that all evidences can be integrated with low computational complexity.

There are various ways to take a decision in the D-S framework. The main decision rules [17] are the maximum of belief, the maximum of plausibility, and the center of the interval whose boundaries are belief and plausibility. Considering the characteristics of the pedestrian detection, this paper chose the belief function as the decision rule.

3. Feature Abstraction

The confirmation of fused features is the first step in D-S fusion framework. The goal of feature computation is to find clues about the presence of pedestrians in video images. The choice of features directly influences the detection performance of the pedestrian detection algorithm. As the proposed detection approach must be generic, we have to find features common to the pedestrians. Therefore, considering the diversity of pedestrian appearances, poses, and behaviors, especially the occlusions and light conditions in the road traffic scenes, motion features would be incorporated into the detection task besides classical shape features. Furthermore, with the aim of pursuing the more efficient pedestrian detection effect, this paper proposed two video-based motion feature descriptors. In this section, the traditional shape and motion features to be implemented in this paper will be briefly introduced, and our proposed motion features and their computation procedure will be described in detail.

3.1. Traditional Features and Performance Analysis

3.1.1. Shape Features. Pedestrians possess many kinds of features, while shape features are the most common and widely used due to their merits of visibility and simplicity. In traffic scenes, the objects with highest occurrence frequency are the vehicles and pedestrians, while there is a big difference between their geometry features. The human body is in the stripe shape while a pedestrian stands or walks; thus compared with vehicles, pedestrians have bigger length-to-width ratio. Similarly, pedestrians have significantly different width and area values from that of vehicles. When the value of the length-to-width ratio is bigger than a certain threshold, or the width value is smaller than a predefined threshold, or the area value is within a predefined threshold interval, we may consider that a pedestrian occurs. In our experiments we used the method presented in [18] to calculate the values of the length-to-width ratio, width, and area of pedestrian candidates.

3.1.2. Motion Features. Although many works implementing shape features to recognize pedestrians have been reported in literature, the pedestrian detectors based on shape features are in many cases not sufficient to achieve satisfactory results in complex weather or scenarios conditions. Because shape features are subject to be straightly affected by the results of pedestrian segmentation generally, it is hard to obtain the satisfactory contour of a pedestrian from traffic images especially under the situations of the complex traffic background, illumination variation, shadow effects, and pedestrian overlapping. For example, pedestrians can be partially occluded by common urban elements, such as parking vehicles or street infrastructure. In such cases, we fail to segment complete pedestrian candidates, resulting in incorrect shape feature values.

To circumvent above problems, motion features are necessary to be considered. Motion features are much more essential and robust in describing pedestrians. Generally, the speed of moving pedestrians is obviously smaller than that of running vehicles. Consequently, the speed metric becomes one basic and typical motion feature which has been heavily studied in many literatures [4, 5, 8, 19]. When the speed value of the candidate target is within a predefined threshold interval, we may consider that a pedestrian occurs. In our experiments, following the definition of pedestrian speed presented in [19], we calculate the speed value of pedestrian candidates using our proposed feature-point-based tracking.

3.2. Proposed Motion Features and Performance Analysis. The extraction procedure of speed feature presented in the literature available always involves the pedestrian segmentation [4, 5, 8, 19], thus inevitably inducing the instability and inaccuracy of pedestrian detection as mentioned above. Therefore, to get around above problems and significantly improve the pedestrian detection performance, we develop two much more discriminative motion features and they are computed based on our proposed feature-point tracking method without involving image segmentation. Our feature-point-based tracking method will be outlined below.

First, we use the Moravec detector [20] to detect corners in the video images, and for the purpose of reducing noise interference we improve the Moravec detector to select high-quality corners for a moving object. The moving object extraction would be done before the corner detection to eliminate unnecessary corners and reduce the calculating amount. We obtain the pedestrian candidate through the accumulating frame difference method [21] followed by the morphological opening and closing operation. Second, considering that pedestrian and vehicle motions can be treated as uniform linear motions between consecutive frames, their corners will be predicted by Kalman filter, which will reduce searching scope and calculating amount. Third, a searching area around the predicted corner in the current frame and a matching template around the corner in the previous frame are established. Then a full search matching is carried out in the searching area with the matching template. Finally, we will get the motion trajectory of every corner, while irregular track trajectories are removed [22]. For details of our proposed tracking procedure, please see our recent works presented in [23].

Therefore, the moving targets are successfully tracked and their trajectories are obtained by our feature-point-based tracking method. Trajectories provide rich information about pedestrian motion states. With the help of behavior analysis based on the trajectories of moving targets, we construct two motion feature descriptors as follows to identify pedestrians.

3.2.1. Trajectory Smoothness. Generally, when pedestrians walk on the road, their walk speeds are always slow and their moving trajectories are usually not smooth (see Figure 1(a)) due to unsteady walking gait. On the contrary, the speeds of vehicles are relatively faster and vehicles almost would not sway, so their moving trajectories are much smoother (see Figure 1(b)). Therefore, trajectory smoothness degree (TSD) can be regarded as one distinct feature of pedestrian incidents. We defined it as the ratio of the length of the arc to that of the chord; that is,

$$\text{TSD} = \frac{\text{Arclength}}{\text{Chordlength}}, \quad (9)$$

where the chord and arc of a certain trajectory are illustrated in Figure 1(c) and X and Y are the horizontal and vertical axis in the image plane, respectively. And

Chordlength

$$= \sqrt{(x(k+n-1) - x(k))^2 + (y(k+n-1) - y(k))^2},$$

Arclength

$$= \sum_{i=0}^{n-1} \sqrt{(x(k+i+1) - x(k+i))^2 + (y(k+i+1) - y(k+i))^2}. \quad (10)$$

$x(\cdot)$ and $y(\cdot)$ are the horizontal and vertical axis values of each corner in the tracking sequence, respectively. We calculate the TSD defined in (9) with the use of n corners after the

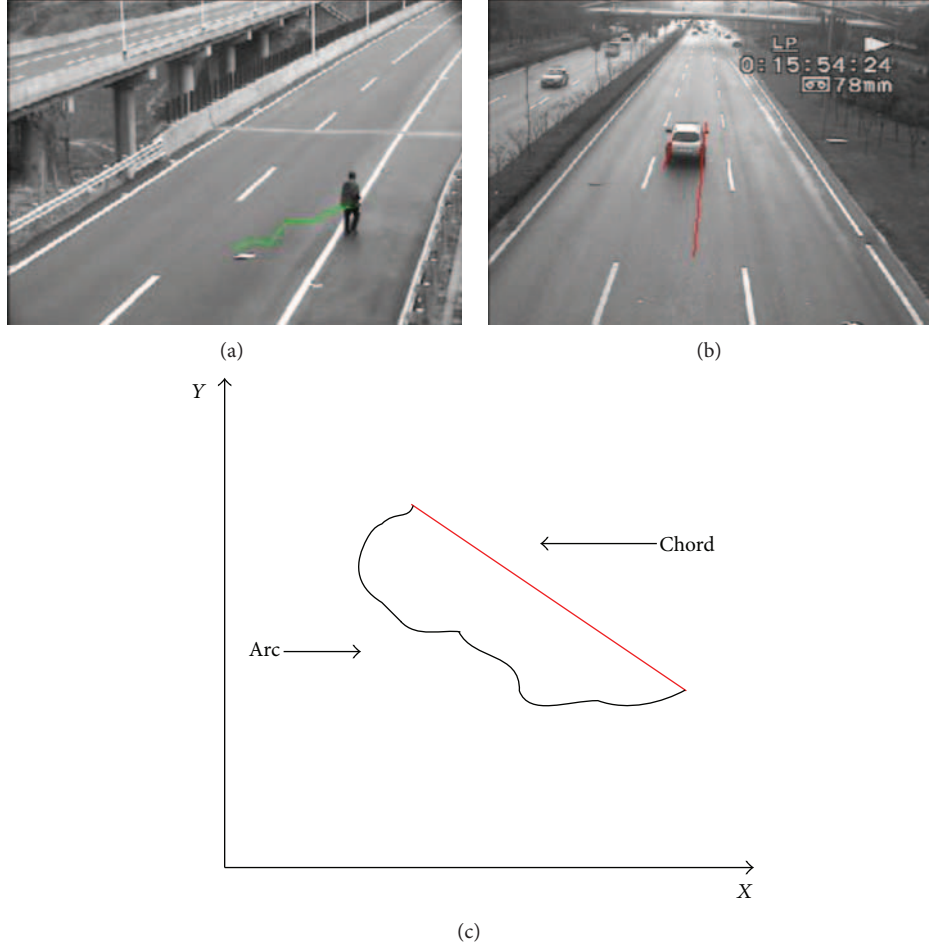


FIGURE 1: Tracking trajectories ((a) pedestrian trajectory; (b) vehicle trajectory; (c) the chord and arc).

k th tracked corner. In experiments we generally assume $n = 30$, and each corner in the tracking sequences is determined using five consecutive frames of monitoring videos. When the TSD is bigger than a presupposed threshold, we identify that the tracked target is a pedestrian.

3.2.2. Trajectory Vector Field. The abnormal targets appearing in the video have a great influence on the normal trajectory vector field. That is, the motion estimation of a pedestrian can be realized by using the overall trajectory vector field. Based on the first M consecutive frames (we always let $M = 1500$ in our experiments), we used the tracking method described in 3.2 to record all trajectories of normally running vehicles on the road and considered each trajectory as a two-dimension vector with the direction and displacement relative to the reference of the camera screen. All the vectors thus form a normal trajectory vector field shown in Figure 2(a). For each curve in the vector field, the distance between every two arrows reflects the current speed of a moving vehicle, and the direction of the arrows indicates its current direction.

Continuing to track the subsequent moving targets, the trajectory vector field updates in real time. In the current trajectory vector field illustrated as Figure 2(b), the green and red curves denote the trajectory vector of the vehicle and

the pedestrian, respectively. Their speed deviation DV and direction deviation $D\theta$ are described in

$$D\theta = \sum_{j=1}^p \theta(j) - \theta_{\text{nor}}(j), \quad (11)$$

$$DV = \sum_{j=1}^p V(j) - V_{\text{nor}}(j), \quad (12)$$

$$\theta(j) = \arctan \left(\frac{y((m-1)j) - y((m-1)(j-1))}{x((m-1)j) - x((m-1)(j-1))} \right), \quad (13)$$

$$\theta_{\text{nor}}(j) = \arctan \left(\frac{y_{\text{nor}}((m-1)j) - y_{\text{nor}}((m-1)(j-1))}{x_{\text{nor}}((m-1)j) - x_{\text{nor}}((m-1)(j-1))} \right), \quad (14)$$

$$V(j) = \left(\left((x((m-1)j) - x((m-1)(j-1)))^2 + (y((m-1)j) - y((m-1)(j-1)))^2 \right)^{1/2} \times \left(\frac{(m-1)}{v_0} \right)^{-1} \right), \quad (15)$$

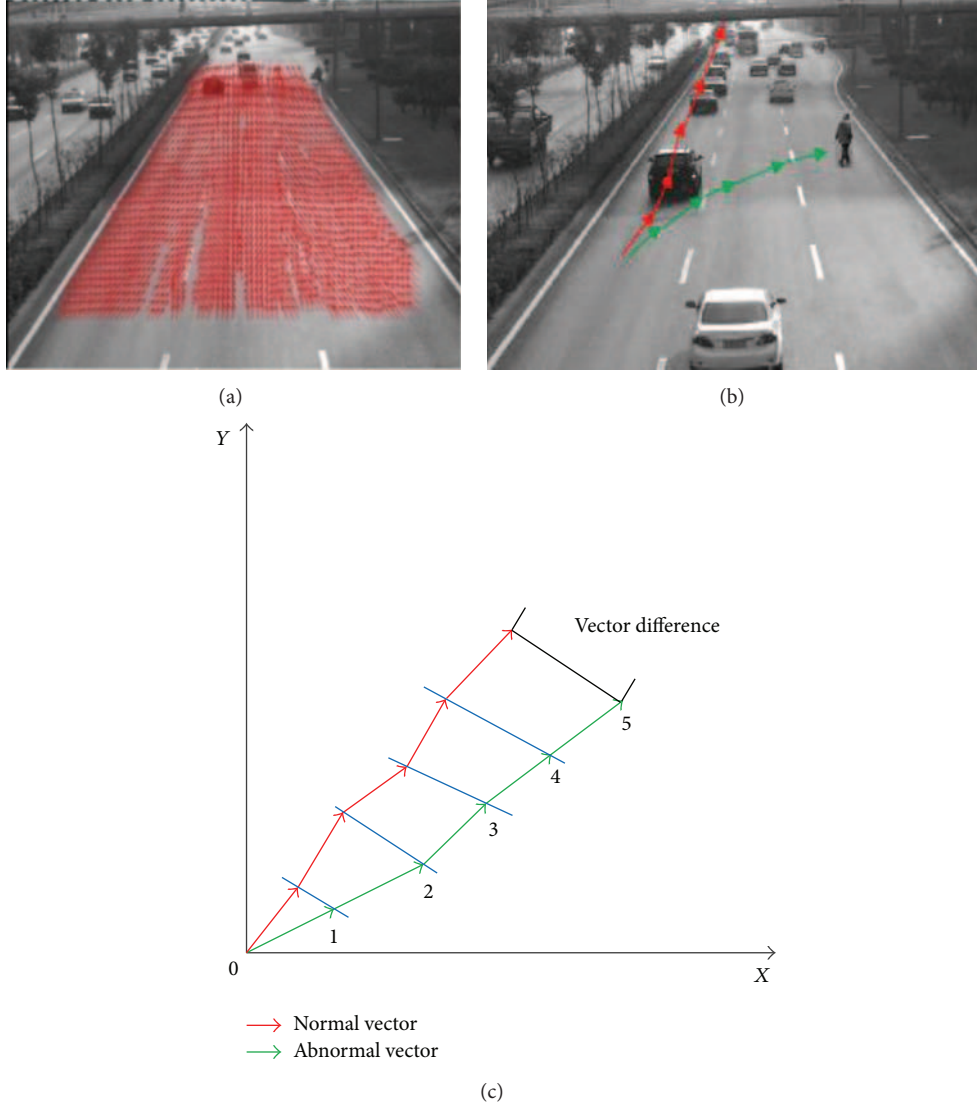


FIGURE 2: Trajectory vector field ((a) normal trajectory vector field; (b) trajectory vector field in case of a pedestrian incident; (c) trajectory vector field analysis).

$$\begin{aligned}
 V_{\text{nor}}(j) &= \left(\left(x_{\text{nor}}((m-1)j) - x_{\text{nor}}((m-1)(j-1)) \right)^2 \right. \\
 &\quad \left. + \left(y_{\text{nor}}((m-1)j) - y_{\text{nor}}((m-1)(j-1)) \right)^2 \right)^{1/2} \\
 &\quad \times \left(\frac{(m-1)}{v_0} \right)^{-1}.
 \end{aligned} \tag{16}$$

$\theta(j)$ represents the direction of the j th trajectory vector of an abnormal target in the trajectory vector field and $\theta_{\text{nor}}(j)$ the direction of the normal target. $V(j)$ and $V_{\text{nor}}(j)$ denote the actual speed and normal speed, respectively. We use p trajectory vectors to calculate the $D\theta$ and DV , while every m tracking corner is fitted into one trajectory vector. $x(\cdot)$ and $y(\cdot)$ ($x_{\text{nor}}(\cdot)$ and $y_{\text{nor}}(\cdot)$) are the horizontal and vertical axis

values of each actual corner (normal corner) in the tracking sequence, respectively. v_0 is the playing-speed of the traffic monitoring video. In our experiments we usually suppose both p and m to be five as illustrated in Figure 2(c) and v_0 to be 25 frames per second. When the values of $D\theta$ and DV are bigger than the presupposed threshold, we recognize that an abnormal target, like a pedestrian, emerges on the road.

3.2.3. Performance Analysis. As well known, the main challenge in the pedestrian detection is that the required performance is quite demanding in terms of speed and robustness. Next we will evaluate our proposed features theoretically in two aspects, that is, robustness and computational complexity.

It is noticed that the previous motion features of pedestrians are usually obtained by tracking method which are based on the segmentation for moving pedestrians [4, 5, 8, 24]. As usual, image segmentation is one of the problems in analyzing

the computer vision, especially under various complex traffic environments, such as cloudy, rainy, and snowy conditions. Many problems such as illumination variations, shadow effects, and overlapping will make the task of segmenting of the individual object difficult [9, 10]. Meanwhile, the complex algorithms of high-level pedestrian modeling and learning create a great amount of computation, which in turn is mostly unstable and time consuming. Different from the idea of tracking objects as a whole, we make an effort to select the most common low-level feature points on pedestrians for tracking without segmentation. Moreover, feature points have the advantage of efficient locating and easy tracking, which is critical for the following target tracking. Accordingly, feature points tracking works well for solving many problems of pedestrian tracking in complex traffic environments and various weather conditions; thus our mined motion features described above can characterize pedestrians with strong robustness.

Furthermore, in our proposed feature point tracking framework, only one or a small number of good or qualified corners are chosen as feature points for a moving object. Therefore, it is quite different from corner point tracking methods available, in which many low-quality corners exist. The causes lie in many aspects. For example, corners may be detected on one pedestrian, and the distribution of corners is usually intensive in a partial region, and virtual corners formed by the background scenes occluded with a pedestrian do not physically exist, and so on. Low-quality corners are unstable during pedestrian tracking process, always leading to tracking errors as well as redundant computations. Consequently our explored motion features may become reliable evidences in making an immediate response to the occurrence of a pedestrian.

4. D-S Fusion Pedestrian Incident Detection Algorithm

4.1. Basic Probability Assignment. Our recognition frame is $\Theta = \{A_1, A_2\}$, where A_1 is the proposition “there is a pedestrian incident,” A_2 the proposition “there is no pedestrian incident,” and $A_3 = (A_1 \cup A_2)$ the doubtful proposition “there is or not a pedestrian incident.” Three measurements, that is, detection correctness rate, false alarm rate, and uncertainty rate, are used to weigh the detection performance of the discussed pedestrian features in this paper [2, 4, 24] and serve as BPA function m .

For multifeature fusion pedestrian detection, the values of BPA function m for each feature must be first confirmed. However, it is hard to get enough samples of pedestrian incidents in realistic road monitoring videos to calculate them. Therefore, this paper uses Monte Carlo simulation [25] with crystal ball [26] to settle sample problems. Monte Carlo simulation allows us to run thousands of iterations of the model quickly and summarizes the entire range of possible outcomes efficiently. It lets us define probability distributions on uncertain model variables and then uses simulation to generate random values from the defined probability ranges.

TABLE 1: Basic probability assignments.

	$m_1(\cdot)$	$m_2(\cdot)$	$m_3(\cdot)$	$m_4(\cdot)$	$m_5(\cdot)$	$m_6(\cdot)$
A_1	0.4974	0.5228	0.5537	0.7221	0.7375	0.7582
A_2	0.4384	0.4143	0.3947	0.2392	0.2304	0.2198
A_3	0.0642	0.0629	0.0516	0.0387	0.0321	0.0220

Before running crystal ball, we need construct an Excel spreadsheet model. Through this model, we obtain the pedestrian detection performance report of the six features j ($j = 1, 2, \dots, 6$) (i.e., length-to-width ratio, width, area, speed, trajectory smoothness degree, and motion vector field). In this model, we suppose three parameters as the triangle distribution. Thus the basic probability assignment $m_j(A_i)$ is realized and displayed in Table 1, where $\sum_{i=1}^3 m_j(A_i) = 1$.

4.2. Fusion Algorithm. The multifeature fusion pedestrian detection algorithm put forward in this paper can be described as follows.

Step 1. Confirm fusion objects; that is, calculate the six features of pedestrian incidents.

Step 2. Determine the detection correctness rate, false alarm rate, and uncertainty rate of each feature for pedestrian detection in the various traffic situations.

Step 3. Obtain the pedestrian detection performance reports with the help of Monte Carlo simulation method (see Section 4.1) to assign the basic credibility.

Step 4. Use evidence combination rule of D-S theory with basic credibility assignment to fuse all features and produce the final credibility;

Step 5. Use the belief function as decision rule to identify whether or not the pedestrian emerges in the traffic scenes.

The proposed pedestrian incident detection approach has the superiority of the high detection performance, strong robustness, low cost, and easy realization.

5. Simulations and Analysis

To test the proposed pedestrian detection approach using D-S multifeature fusion and to evaluate our developed motion features, a large number of simulations and experiments have been carried out under various traffic scenes, and we also compared it with the SVM feature fusion method presented in [14] in terms of the pedestrian detection performance. We compared the number of the traffic events automatically detected (corresponding to the “Detection” in Table 3) by detection algorithms with the number of events detected by human sight (corresponding to the “Inspection” in Table 3) for the same sequence.

5.1. Simulation Experiments. First, we tested our pedestrian detection algorithm offline in terms of its ability to detect

TABLE 2: Pedestrian detection results based on real traffic surveillance data.

Features	$m(A_1)$	$m(A_2)$	$m(A_3)$	Decision results
Length-to-width ratio	0.4974	0.4384	0.0642	A_3
Width	0.5228	0.4143	0.0629	A_3
Area	0.5537	0.3947	0.0516	A_1
Speed	0.7221	0.2392	0.0387	A_1
SVM fusion in [14]	0.8054	0.1937	0.0009	A_1
D-S fusion	0.8357	0.1640	0.0003	A_1
Trajectory smoothness	0.7375	0.2304	0.0321	A_1
Trajectory vector field	0.7582	0.2198	0.0220	A_1
SVM fusion in [14]	0.9213	0.0782	0.0005	A_1
D-S fusion	0.9735	0.0264	0.00012	A_1

TABLE 3: Pedestrian detection results in real traffic scenes.

SVM fusion method with all features extracted in Section 3					
Road number	Detection count	Inspection count	False alarm count	Detection positive rate	Detection time (S)
1	7	6	1/6	6/6	
2	27	24	1/24	22/24	
3	2	2	0	2/2	
4	19	16	1/16	13/16	
5	76	67	4/67	61/67	
6	47	43	3/43	42/43	
Average			7.87%	90.41%	<10
D-S fusion method with all features extracted in Section 3					
Road number	Detection count	Inspection count	False alarm rate	Detection positive rate	Detection time
1	6	6	0/6	6/6	
2	25	24	1/24	23/24	
3	2	2	0	2/2	
4	18	16	1/16	15/16	
5	72	67	3/67	65/67	
6	46	43	2/43	43/43	
Average			4.9%	95.53%	<5

pedestrian events. We prepared many scenarios with different pedestrian events and in different road sections in tunnels and in cities of Xi'an, Shanghai, and Chongqing in China with various conditions, including sunny days, cloudy days, and rainy day and nighttime. The size of each image is 720×288 , and the sampling frequency is 25 fps. The proposed system is implemented with Visual C++ on a raw video format.

Figure 3 shows the offline test scenarios, including Chongqing highway in sunny day and Shanghai highway in sunny day, cloudy day, and at night, and urban road sections in Xi'an and Huaying tunnels. The simulation results of pedestrian event detection in different cases are summarized in Table 2. Line 6 and Line 10 display the values of detection correctness rate, false alarm rate, and uncertainty rate obtained by the SVM fusion method put forward in [14] with the traditional features (please refer to Section 3.1) and with additional our proposed motion features (please refer to Section 3.2). Line 7 and Line 11 show the corresponding values obtained by our proposed D-S fusion method. And other values in other lines are basic probability assignments of single feature same to Table 1.

Comparing Line 7 with Line 6 and Line 11 with Line 10 in Table 2, we can see our proposed D-S fusion approach yields higher detection correctness rate and lower uncertainty rate than the SVM fusion method with the same features. That demonstrates the superiority of our D-S fusion framework over SVM fusion approach in the fact that the appealing D-S evidence theory can deal with uncertainty more efficiently. In addition, the comparison of the Line 10 with Line 6 and Line 11 with Line 7 in Table 2 reveals that using our developed features can improve detection performance for both the D-S fusion framework and SVM fusion approach, implying our proposed two features presented in Section 3 are quite feasible and valuable for pedestrian incident detection. Moreover, during our simulation experiments, we also compared the operation time of our propose D-S fusion pedestrian detection method with that SVM fusion pedestrian method. The records show the former can respond within 2 seconds, much quicker than the latter of 7 s. SVM fusion approach involves learning procedure on large training data set, leading to excessively great computational complexity. This series of experiments on real traffic surveillance data indicate that the



FIGURE 3: Detection results in various scenarios under different conditions ((a) Chongqing highway, sunny day; (b) urban road in Xi'an, sunny day; (c) Shanghai highway, sunny day; (d) Shanghai highway, cloudy day; (e) Huaying Mountain Tunnel, (f) Shanghai Highway, nighttime).

proposed pedestrian detection method works well with high detection performance for various weather conditions.

5.2. Online Experiments in Real Traffic Scenes. Our pedestrian detection method was tested online on six urban roads in Shanghai to see whether the system could accurately detect the pedestrian events under various conditions, including sunny days, cloudy days, rainy day, and nighttime. We have implemented the surveillance system with a camera network including six cameras. The heavy computational operations of feature point extracting and tracking as well as pedestrian motion feature extraction are placed on the video processors, while the detection result and relevant data are saved on the host PC. The video processor is composed of 20 DSPs, and

each DSP processes the data collected by one camera. Meanwhile, the department of traffic management can monitor real-time traffic conditions through the Ethernet.

Our surveillance system worked 24 hours from November 2, 2013, to December 2, 2013, including sunny day, cloudy day, and rainy day. And different test roads have different traffic complexity and different pedestrian occurrence probability. The testing results obtained by our method and SVM-based fusion method are listed in Table 3, where "Average" denotes the average detection positive rate, average false alarm rate, and average response time of all detection results on six roads. As displayed in Table 3, we can see that the proposed method has the superiority with higher detection positive rate and lower false alarm rate than the SVM-based fusion method in the real traffic scenes. Additionally, from

Table 3 we know the pedestrian detection times for SVM-based fusion method are less than 10 s, whereas the proposed method needs 5 s at most. Therefore, the proposed method would have better performance in real-time applications.

6. Conclusions

D-S theory is powerful to deal with uncertain information, and traffic monitoring videos include rich information of pedestrian incidents. So taking advantages of them, we constructed a multifeature D-S fusion pedestrian detection approach with the important features abstracted from traffic videos. Furthermore, besides the traditional shape features, we developed two efficient motion features for pedestrian incidents by employing an improved feature-point tracking framework without object segmentation. Theoretical analysis and experiment results on real data and in real traffic scenes consistently show our explored two features are effective and our proposed fusion approach has superior pedestrian detection performance under various weather conditions. Theoretical analysis and experiment results also consistently show that the proposed method outperforms SVM-based multifeature fusion approach for pedestrian detection in terms of recognition ability and robustness and response time in various real traffic scenes.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors acknowledge the support of the National High-Tech Research and Development Program (863program) of China (no. 2012AA112312). They furthermore thank the anonymous reviewers for valuable comments.

References

- [1] Z. Lin and L. S. Davis, "Shape-based human detection and segmentation via hierarchical part-template matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 604–618, 2010.
- [2] S. Munder, C. Schnorr, and D. M. Gavrila, "Pedestrian detection and tracking using a mixture of view-based shape-texture models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 333–343, 2008.
- [3] F. H. C. Tivive and A. Bouzerdoum, "A biologically inspired visual pedestrian detection system," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '08)*, pp. 703–709, Hong Kong, June 2008.
- [4] M. Enzweiler, P. Kanter, and D. M. Gavrila, "Monocular pedestrian recognition using motion parallax," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '08)*, pp. 792–797, Eindhoven, The Netherlands, June 2008.
- [5] Y. Chen, Q. Wu, and X. He, "Motion based pedestrian recognition," in *Proceedings of the Congress on Image and Signal Processing*, 2008.
- [6] M. Bertozzi, A. Broggi, M. del Rose, M. Felisa, A. Rakotomamonjy, and F. Suard, "A pedestrian detector using histograms of oriented gradients and a support vector machine classifier," in *Proceedings of the IEEE Intelligent Transportation Systems Conference*, pp. 143–148, Seattle, Wash, USA, October 2007.
- [7] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.
- [8] G. Gate and F. Nashashibi, "Using targets appearance to improve pedestrian classification with a laser scanner," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 571–576, Eindhoven, The Netherlands, June 2008.
- [9] N. Buch, S. A. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 920–939, 2011.
- [10] J. M. A. Alvarez and A. M. Lopez, "Road detection based on illuminant invariance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 184–193, 2011.
- [11] E. Waltz and J. Llinas, *Multisensor Data Fusion*, Artech House, Boston, Mass, USA, 1990.
- [12] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "A framework for decision fusion in image forensics based on Dempster-Shafer theory of evidence," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 593–607, 2013.
- [13] Y. Matsui, M. Hitosugi, T. Doi, S. Oikawa, K. Takahashi, and K. Ando, "Features of pedestrian behavior in car-to-pedestrian contact situations in near-miss incidents in Japan," *Traffic Injury Prevention*, vol. 14, no. 1, pp. 58–63, 2013.
- [14] A. Pérez Grassi, V. Frolov, and F. Puente León, "Information fusion to detect and classify pedestrians using invariant features," *Information Fusion*, vol. 12, no. 4, pp. 284–292, 2011.
- [15] N.-E. El Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: progress and challenges—a survey," *Information Fusion*, vol. 12, no. 1, pp. 4–10, 2011.
- [16] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [17] C. K. Murphy, "Combining belief functions when evidence conflicts," *Decision Support Systems*, vol. 29, no. 1, pp. 1–9, 2000.
- [18] L. Zhao and C. E. Thorpe, "Stereo and neural network-based pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 3, pp. 148–154, 2000.
- [19] S. Milch and M. Behrens, *Pedestrian Detection with Radar and Computer Vision*, The Pennsylvania State University, 2001.
- [20] H. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover," Tech. Rep. CMU-RI-TR-80-03, The Robotics Institute, 1980.
- [21] J. Zhou, D. Gao, and D. Zhang, "Moving vehicle detection for automatic traffic monitoring," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 51–59, 2007.
- [22] B. Johansson, J. Wiklund, P.-E. Forssén, and G. Granlund, "Combining shadow detection and simulation for estimation of vehicle size and position," *Pattern Recognition Letters*, vol. 30, no. 8, pp. 751–759, 2009.
- [23] H.-S. Song, S.-N. Lu, and X. Ma, "Vehicle behavior analysis using target motion trajectories," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3580–3591, 2014.
- [24] D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1239–1258, 2010.

- [25] W. Smythe, C. S. Merle, R. Rustomjee et al., "Evaluation of initial and steady-state gatifloxacin pharmacokinetics and dose in pulmonary tuberculosis patients by using Monte Carlo simulations," *Antimicrobial Agents and Chemotherapy*, vol. 57, no. 9, pp. 4164–4171, 2013.
- [26] I. Shiers, "Looking into the crystal ball," *Commercial Motor*, vol. 19, no. 4, pp. 40–41, 2013.