2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Pedestrain Detection from Motion

A spatial-temporal approach based on walking actions

Mehmet Kilicarslan, Jiang Yu Zheng, Kongstantino Raptis
Dept. of Computer Science
Indiana University Purdue University Indianapolis
Indianapolis, USA
mkilicar@iupui.edu, jzheng@iupui.edu

***Abstract* — Pedestrian detection is a challenging problem studied over decades. Most algorithms are based on human appearance. Only few works consider motion as a feature component. In this paper, however, we tackle this problem only considering short periods of pedestrian walking. This motion does not depend on the variations of pedestrian pose, body shape, illumination, and background. We model pedestrian motion that has unique properties compare to background and rigid objects motion in spatial-temporal motion profiles. This observation helps us to identify pedestrian leg motion along with body motion over a short time period. Our method also works for a vehicle borne camera where background also moves. We achieved more robust results by dealing with crowds, and other degenerating cases of human motion against background and dynamic scenes. The method has a low computational cost on a motion profile and it can be combined with a shape-based method as pre-screening for reducing the false positives. It also provides a feasible way to find human behaviors.**

Keywords—pedestrian motion; pedestrian detection; spatial-temporal filtering; driving video; tracking

I. INTRODUCTION

Pedestrian detection [2,3,4,5,6] is crucial for intelligent vehicle, surveillance, and human behavior understanding. Pedestrian detection in driving video [7] is more challenging than in surveillance video taken by a static camera due to the dynamically changing environment, which prevents background subtraction as detection methods. Many works have explored this topic and achieved great success [8,10] based on identifying human appearances, though such methods may be influenced by variations of pedestrians, complex background, crowds [1,23, 25], and occlusions of other objects [13]. HOG [9,11], Haar-type feature, LBP, and their variations are the main features used in sliding window based shape approaches [14,15].

Motion is another clue not thoroughly explored for pedestrian detection. As initial explorations, either optical flow combined with shape [16] or longer term action in trajectories [17,19] are examined independently. The former added velocity in the recognition if the camera itself has no movement, and the latter employs acceleration, i.e., leg stopping and stepping in walking, even camera is moving. Walking action is common for pedestrians [20,21,26], which is less varied than shape from person to person and is much more different from other rigid motion such as vehicle motorized motion. Either static background or dynamic vehicles captured by a vehicle borne camera has the smooth motion, which is guaranteed by the driving mechanism of four-wheeled vehicles. If the motion can be profiled properly in a spatial-temporal image [18], the pedestrian leg traces can be observed clearly as a chain of rings. In our previous work on pedestrian trajectories, HOG and template matching are combined to obtain leg crossings on pedestrian trajectories (HOGTM) [17]. However, such ideal templates are unable to handle the situations of crowds with overlapped trajectories, and the fast vehicle/camera motion that deforms trajectories at rings and chains. Another approach in [19,24] finds non-smooth motion flow (NSM) at legs and body that appear as corner points on walking trajectories in contrast to the smooth background and vehicle traces. It has higher detection rate on pedestrians including crowds, but also brings a high false positive rate from intersecting traces due to occlusion, and from the aliasing and digital noise due to the insufficient temporal sampling of video on fast passing objects.

The objective of this work is to achieve pedestrian detection robustly from motion, which includes more cases of vehicle/camera motion and pedestrian depth, crowds, and occlusion with other object motion/traces. We design spatial-temporal filters for detecting pedestrian trajectories at signature spots in the motion profiles, which turns the temporal feature extraction to shape filtering on the motion profiles. The accuracy using public dataset achieves good results.

This paper provides a comprehensive solution for motion based pedestrian detection. Pedestrians have articulation motion particularly in leg stepping, which is unique in the second derivative of leg positions. This is distinguishable from other object motion with smoothly changed velocities. In the motion profile, pedestrian steps form a chain trajectory. This allows the pedestrian recognition regardless of the variations of pedestrian pose, shape, color, illumination, background, etc.

In the following sections, the paper starts from a sensing scheme to capture motion on road when a vehicle borne camera is moving on street. Then, the motion behavior of pedestrian, background, and possible dynamic vehicles are compared in Sec. II. The spatial-temporal filtering algorithm for the leg chain detection is introduced in Sec. III. Section IV is the pedestrian recognition from the leg information. Section V discusses the experimental results and evaluation, followed with conclusion in Section VI.

II. OBTAINING MOTION IN DRIVING VIDEO PROFILES

A. Sensing Scenes at Different Depths

To profile motion in a visual space for analysis, we first

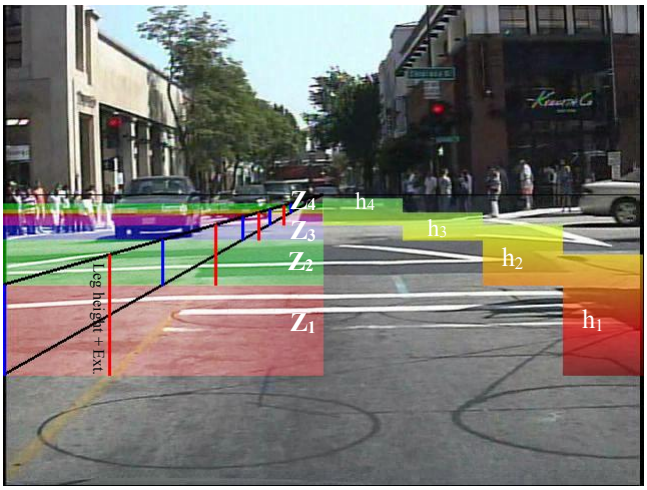


Fig. 1. Image zones (colored on right) are located under the horizon to cover legs at different depths (colored on left) ahead. Displayed zones only indicate their heights while the horizontal span is the entire frame.

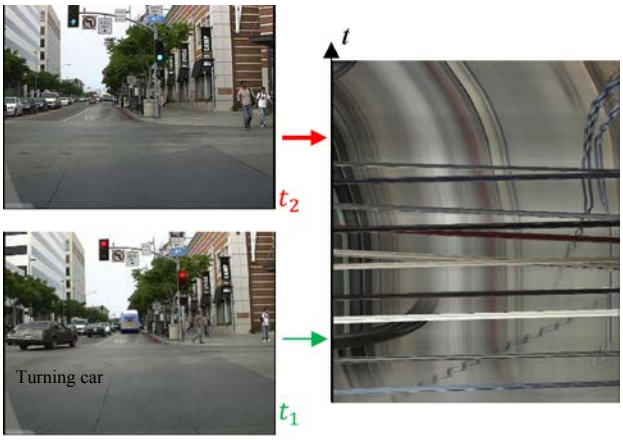


Fig. 2. A motion profile $P(x,t)$ (right) generated from a video clip (left) at a high position near the horizon. Two frames at arrow pointed times are displayed. Horizontal traces in the motion profile are passing vehicles and vertical curves are background motion. Leg motion of pedestrains is also visible as chains.

create spatial-temporal images at different viewing angles. A forward-looking video camera is set near the back mirror of a car during vehicle motion. The horizon is located once at the height of FOE (Focus of Expansion) in the frame, which is obtained via the color accumulation of frames over a section of video when the car moves straight. Below the horizon, several horizontal zones, e.g., four zones z_1~z_4 are set in the frame to cover depth ranges 5-15m, 15-30m, 30-60m, and beyond 60m briefly on the ground. According to the projected positions of the depth ranges in the frame, as well as the average leg height of pedestrians, we set the zones at coordinates $y_1<y_2<y_3<y_4$ from bottom to horizon in the video frame. Their heights are set $h_1>h_2>h_3>h_4$, with h_1 at lower zone for close pedestrians and h_4 at upper zone for far depth, respectively. These zones have overlaps under the perspective projection. With the fixed camera height close to the height of pedestrians, these zones can cover most part of pedestrians in all distances. In the upper zone, background objects, passing vehicles, body of close pedestrians, legs and body of far pedestrians are captured. In the lower zone, legs of close pedestrians and road surface are captured. Other zones include mixtures of these two cases.

Multiple motion profiles, e.g., $P_i(x,t)$, i=1,..4, are condensed from these zones in this work. We employ the temporal profile method in [18] to sample all the zones. In detail, pixels in a zone are vertically averaged to obtain a pixel line, and lines from consecutive frames are connected to form a motion profile image, $P(x,t)$. As shown in Fig. 2, the profile contains motion trajectories of scenes and pedestrians, referred to *traces*, for analyzing long-term motion than the between-frame *optical flow*. From four zones, therefore, four stacked motion profiles are displayed in Fig. 3. We can see some close-to-horizontal trajectories from cross-street walking and the waved background motion vertically. One of the most complex motion profile from a crowd in Caltech database [2] is also shown in Fig. 4. We search pedestrian walking traces in the limited number of motion profiles such that the window scanning over the entire frame is unnecessary. On the other hand, the zone heights h_i should not be too large so as to prevent the leg traces being blurred out in the motion profiles due to the color condensing. If the camera is set at a lower position on a car, which is a preferred case, zones will be more overlapped in the frame, and the horizontal lines of sight can reach a farther depth. In such a situation, fewer zones can cover entire depth range. Because the motion profiles are from overlapped zones, a walking pedestrian will be captured at least in one profile. Detecting a pedestrian trace $x_i(t)$ in i^{th} motion profile will flag the zone at frame t as pedestrian occupied.

B. Leg Chains of Pedestrian Traces and Object Traces

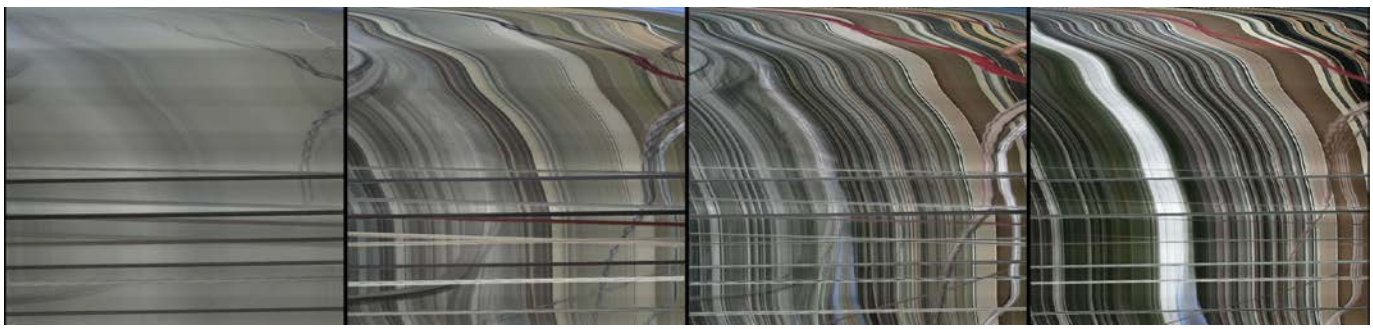


Fig. 3. Motion profiles $P_i(x,t)$ from zones z_1, z_2, z_3, and z_4 in the order of low-to-high. The vertical axes are the time/frame, and the horizontal axes are image x coordinates. Slanted and close-to-horizontal traces are crossing vehicles in front of camera. The lower motion profile (left) covers more flat ground. Upper profile (right) shows scene traces projected from far depth. Pedestrian trajectories are visible as non-smooth chains near the ground on the right margin.

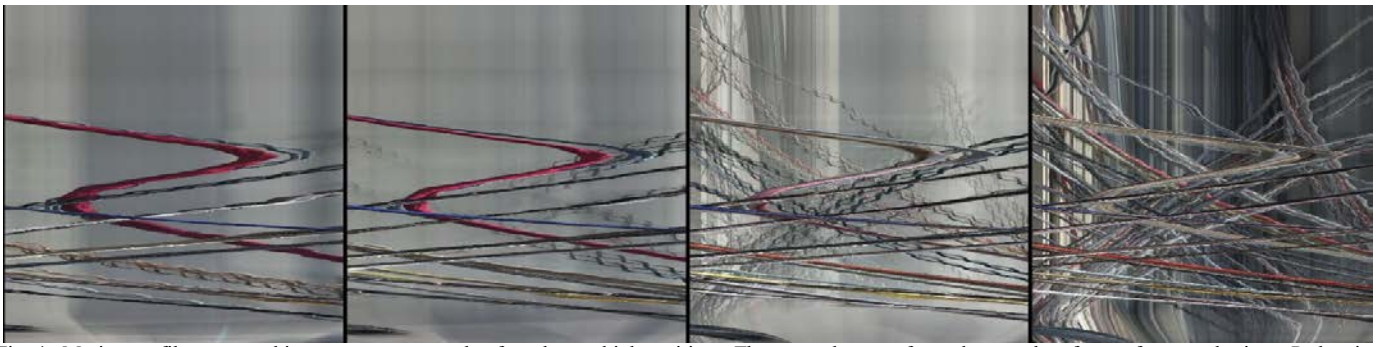


Fig. 4. Motion profiles averaged in zones z_1, z_2, z_3, and z_4 from low to high positions. They record scenes from close road surface to far-away horizon. Pedestrian crowd moves in front of the camera, which makes complex trajectories. The time axes are all upward.

Table I Motion characteristics between pedestrian and other object traces in the motion profiles

Traces	Spatial property: image width		Spatio-temporal property: velocity		Temporal property: frames	
	Edge detection	Leg-width stripe detection	Trace slanted (fast)	Trace vertical (slow)	Short stripe	Long traces
Objects occluding others	**Appear as edges**	Narrow poles and fence are stripes	**Moving fast due to close depth**			**Long and continuous**
Background Objects	**Mostly edges**	Some surface patterns may appear as stripe	Skew only during vehicle rotation	**Far objects have slow motion**	Occluded case	**Generally, long**
Dynamic vehicles	**Mostly edges**	Partially appear as stripe, e.g., tire	Passing vehicle has fast speed	Parallel ones are slow relative to camera	Passing one is short time	Parallel one is long time
Stopping legs	Edges on two sides	**Leg bounded by two edges**		**Close to vertical even skewed**	**Within a step cycle**	
Stepping legs	Unstable edge/ motion blur	**Leg has response**	**Fast and thus slanted in profile**		**Even shorter**	

There are distinct motion patterns for pedestrian among the scene background. More detailed than the pedestrian trajectories analyzed in [17, 19], we found the properties in the motion profile (Fig. 5) under the ideal case of slow vehicle/camera motion:

(a) Leg chain structure: two legs move alternatively with stepping and standing periods. Pedestrian trajectory is visually formed in the motion profile as a chain composed of rings with leg crossing (X-pattern) and starting/ending points (L-patterns).

(b) Temporal property: the period of a walking cycle is 1.26 second in average [22] irrelevant to the vehicle/camera moving speed. Short-term standing leg has an average time of one second.

(c) Spatial property: the spatial interval (horizontal width) of steps in the motion profile is related to real step distance and pedestrian depth in the 3D space.

In the real situations, the pedestrian leg chain undergoes several types of deformation that previous HOGTM method [17] is incapable to deal with.

(d) Chain degeneration: when the vehicle/camera has a fast translation or turning, the image velocity of scenes increases. This will skew the leg chain of pedestrian trace more or less horizontally in the motion profile, making the leg crossing degenerated as in Fig. 6.a. The HOGTM method [17] based on nice leg chains will have a high false negative rate.

(e) Torso are mixed with arms in the pedestrian traces (Fig. 6.b). Some pedestrian wearing skirt also affects the leg chain detection.

Although object trajectories in video are mostly smooth in long term motion, front narrow objects may occlude background objects. Their trajectories form T-junctions in the motion profile as shown in Fig. 4. Through the examination of a large volume of videos and their motion profiles, we summarize detailed relation between various traces and their responses to features such as edge, stripe, direction, and length in Table I.

t

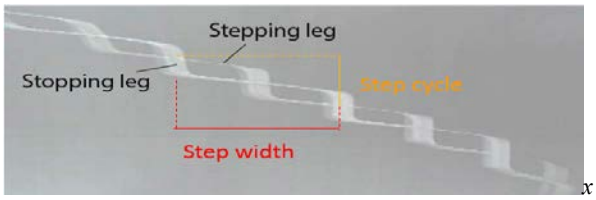


Fig. 5. Typical leg chain of pedestrian walking in the motion profile with consistent step width and period.

III. Leg Trajectory Detection by Filtering

The pedestrian detection now becomes the recognition of leg-chains with rings and crossings in the motion profile, which

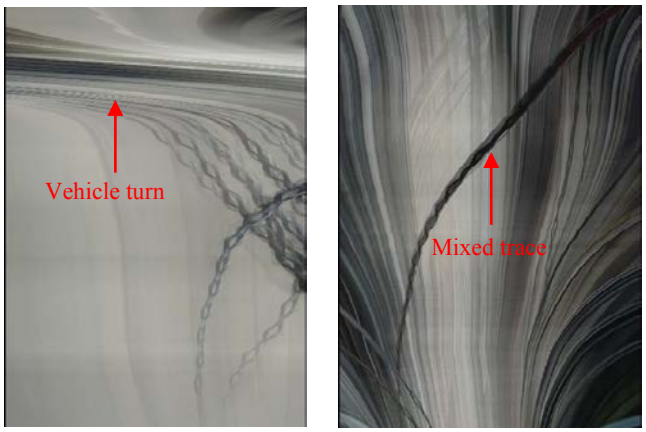


Fig. 6. Degenerate traces of leg chains in the motion profiles due to the skew of the motion profile during vehicle turning. (left) Trace skewed left when the car turned right. (right) Body trace mixed with leg trace after profiling.

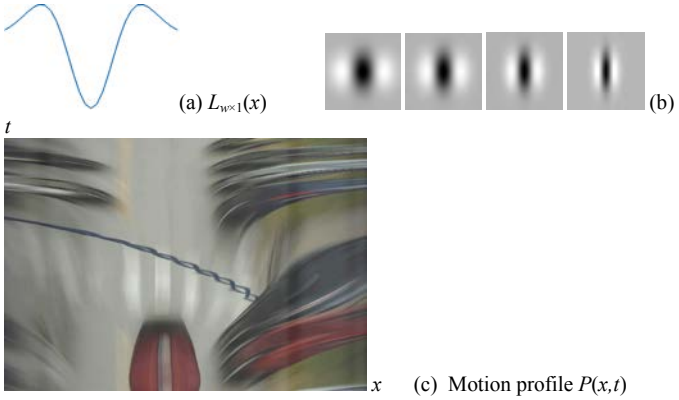


Fig. 7. Filters to obtain leg-width stripes in the motion profiles. (a) Profile of a filter to extract leg-width stripes. (b) Different widths of filter for different zones from low-to-high to obtain leg traces at different depths. (c) A motion profile with pedestrian leg trace at center along with car traces and ground patterns. (d) Filter response in both positive and negative values. (e) Zero-crossing points (i.e., edges) bound stripes. Blue indicates positive values and green indicates negative values in $S(x,t)$.

is easy by humans. If a part of chain is identified at any zone, we claim the human is detected there. With a smooth camera motion, background motion has long and smooth traces in contrast to rhythmic pedestrian walking. To locate general leg chains including degenerate cases, we focus on a signature pattern in this work, i.e., thick and short *strokes* of stopping (standing) legs on pedestrian traces, because thin and fast traces of stepping legs may be extended largely in width if they are skewed spatially.

We classify traces to pedestrian and non-pedestrian according to their spatial properties, temporal continuity and smoothness. Such properties can be described spatially by {edge, stripe}, spatio-temporally by {slanted, vertical} trace orientation, and temporally by {long, short} trace period in the motion profile. As indicated in Table I, a pedestrian leg has an estimated width and thus appears as a *stripe* bounded by two *edges* in the motion profile. The orientation of a trace is slanted if the motion is fast. An object trace is temporally long and smooth, while walking legs alternate frequently. Therefore, a pedestrian chain is distinguishable from its short standing period and leg crossing in the motion profile.

In practice, we use three concatenated filters to (1) spatially locate leg-width *stripes* among all traces, (2) spatio-temporally identify slow-moving stripes, and then (3) temporally track the stripe length for short strokes. This is more scalable and robust to deformed traces than previous HOGTM and corner based methods [17,19].

Spatial filtering: the minimum clue on shape used in this work is the average leg-width w in a frame/profile from a particular depth after brief calibration. We use a 1D horizontal Laplacian of Gaussian filter, $L_{w\times 1}(x)$, as in Fig. 7a to obtain leg-width stripes in the motion profile. That is

$$S(x,t) = L_{w\times 1}(x) \oplus P(x,t) \qquad (1)$$

where $\oplus$ indicates convolution. One example resulted is in Fig. 7d containing both standing and stepping leg traces after this filtering. However, strong edge traces also have low responses to the filter, which have to be removed. To confirm a stripe other than an edge, we search for two *zero-crossing* points, i.e., edges, around the peak response of $S(x,t)$. Or equivalently, search for two neighboring peaks with the opposite sign from the focused peak in $S(x,t)$ (Fig. 7e).

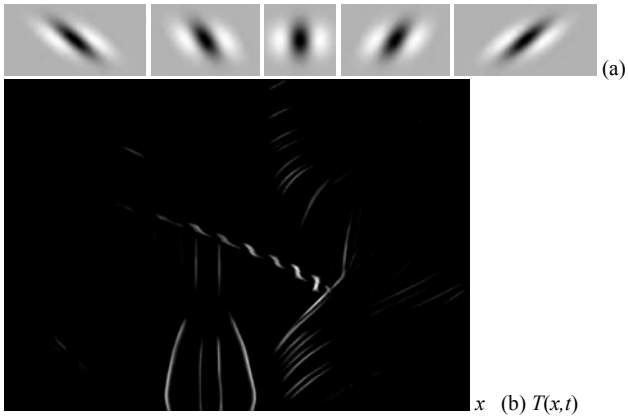


Fig. 8. Spatio-temporal filter results to remove fast traces. (a) Five skewed orientations of filter $R_\theta(t)$ to tolerate varied leg-motion speed. Composite filter effect is shown. (b) By selecting the maximum response from filters in 5 directions, the slow-moving traces close to vertical is remained.

Spatial-temporal filtering: the traces in $S(x,t)$ are then examined spatio-temporally by their orientations. The standing legs have slow image motion in the motion profile even being skewed spatially by fast camera motion. By skewing a vertical 1D temporal averaging filter $R_{1\times m}(t)$ spatially in the x direction, an oriented filter, $R_\theta(t)$, $\theta\in[-45°,45°]$ is generated for the motion profile. Here m is consistent with the average leg-stopping period of 1 second. We identify orientated traces of possible pedestrians using five directional Gaussian filters, i.e.,

$$T(x,t) = \max_{\theta_i}\{R_{\theta_i}(t) \oplus S(x,t)\}, \qquad i = 0, \pm 1, \pm 2 \qquad (2)$$

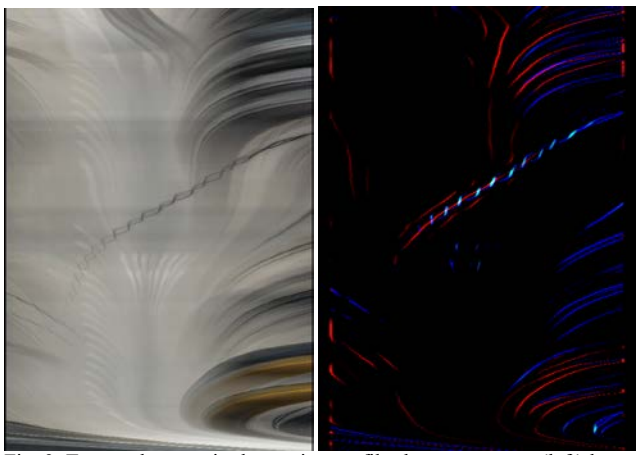


Fig. 9. Trace coherence in the motion profiles between zones. (left) lower profile, (right) Detected stopping legs are connected with stepping leg traces from the same zone and overlapped with body traces from higher zone. Red channel shows possible body traces at upper motion profile, green channel is for standing leg strokes, and blue channel is stepping leg traces.

The composite effect of filters L and R is given in Fig. 8a. The resulting $T(x,t)$ removes fast object traces but keep long and slow object traces as in Fig. 8b. A threshold δ_1 is set to select stripes that have high responses in $T(x,t)$ such that fast/slanted traces are ignored.

Temporal filtering: now the trace continuity is examined through the tracking of stripes along their peaks in $T(x,t)$. The trace length is counted incrementally during tracking. Only the strokes finished shorter than a threshold δ_2 are output as candidates of stopping leg trace. Most object traces have been removed at this stage, except some short traces cut by other traces due to occlusion and fast vehicle turning.

IV. Pedestrian Detection By Motion Continuity

In the first stage, we have detected pedestrian walking trajectories in the profile. However, object motion may also be misclassified if only short strokes are targeted, which yields false positives. The special events causing these are: (1) occlusion, i.e., slow moving object traces are frequently occluded by slanted traces of fast moving vehicles. Such short stripes form T-junctions in between slanted traces shown in Fig. 2. (2) Sudden turning of vehicle/camera, may cause short segments on all traces that have close-to-zero velocities shortly (trace direction changing from – to + or vice versa during vehicle turning). These segments are not linked but kept in parallel. (3) Some surface marks on road and tire traces also remain as false positives.

In the second stage, we use the motion continuity of two legs in walking cycle to remove false positives. After tracking for short strokes in the previous section, we track the body traces to link the stopping strokes and thus reject standalone short stripes.

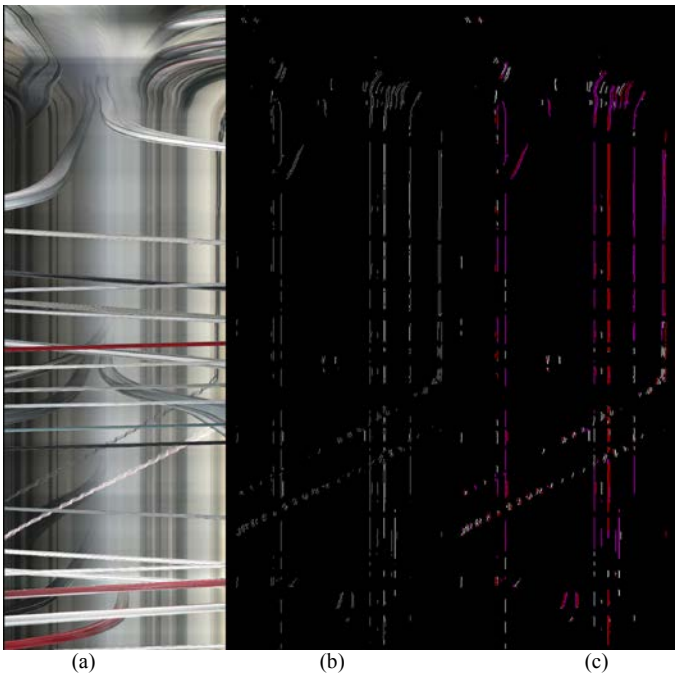
Stepping leg traces can also be detected by the first 1D filter of Laplacian of Gaussian $L_{w\times 1}(x)$, and they are tracked along edges. If standing leg strokes are connected with traces of stepping legs, leg traces are completed as a chain. We do this extension for the result visualization as well. Also, after locating leg traces at a profile, the trace in the higher profile may contain the body trace. The body trace can be marked as pedestrian and this also confirm the leg traces below in the lower zone. Generally, close pedestrian bodies are captured as non-smooth traces in a higher zone. The overlap of traces between upper and lower profiles show the propagation of pedestrian *id* upward. Figure 9 is an example that overlap stopping and stepping leg traces for the confirmation of pedestrians.

V. Experiments

The method has been tested on Caltech video database [2] and partially on TASI Driving Video [22]. The former has a higher camera position and the image is 640×480 pixels, while the latter has images of 1280×720 pixels in a wider field of view. The horizon is calibrated for Caltech data set. The real parameters for setting zones in Caltech video are summarized in Table II. Only two zones are set for TASI data set for the lower camera positions. We select trace orientation θ to be five directions among [-45°, 45°], and the temporal m is 15 pixels/frames in the motion profile over 24~30 frames of standing legs. Tracking threshold δ_2 for short strokes are between (10-45) frames. Figure 10 gives an intermediate results marked as stripe and edges along with their lengths.

Table II: Setting parameters for implementation of filtering

Zones from bottom	1	2	3	4
y coordinates in frame	286 370	254 310	233 270	218 243
Height h_l in pixels	84	56	37	25
Filter width for legs in pixels	45	36	27	21



(a) (b) (c)

Fig. 10 Traces of edges (red), long stripes (pink) and short strips (white) in the results (c) obtained from the filtering result in (b) for the motion profile in (a).

The detected position of a pedestrian is finer than a bounding box horizontally used in traditional human detection. Filtered results of stopping legs are obtained at the center of legs, and the tracked body traces also give continuous existence of pedestrians. We verify the pedestrian results against the ground truth of Caltech database by plotting the one-pixel wide trace onto the trajectories of bounding box with the changing width. Vertically, the detecting results are counted in zones and are less accurate than bounding box. On the other hand, the aspect ratio of pedestrians does not matter in the recognition. We count the number of frames a pedestrian is detected, which is the trace length in the motion profile, in evaluating the detection accuracy of our method. For example, the most complex video in Caltech dataset is a crowd at an intersection (Fig. 4) with our result for 1min clip (1811 frames) in Fig. 11. It results in 285 correct strokes (stopping steps), 88 missed strokes, 28 false positive strokes, and the stroke length is 30 frames in average. In general, the results from all layers are combined by OR operation; if one zone detects pedestrian legs, the above zones are considered as detected.

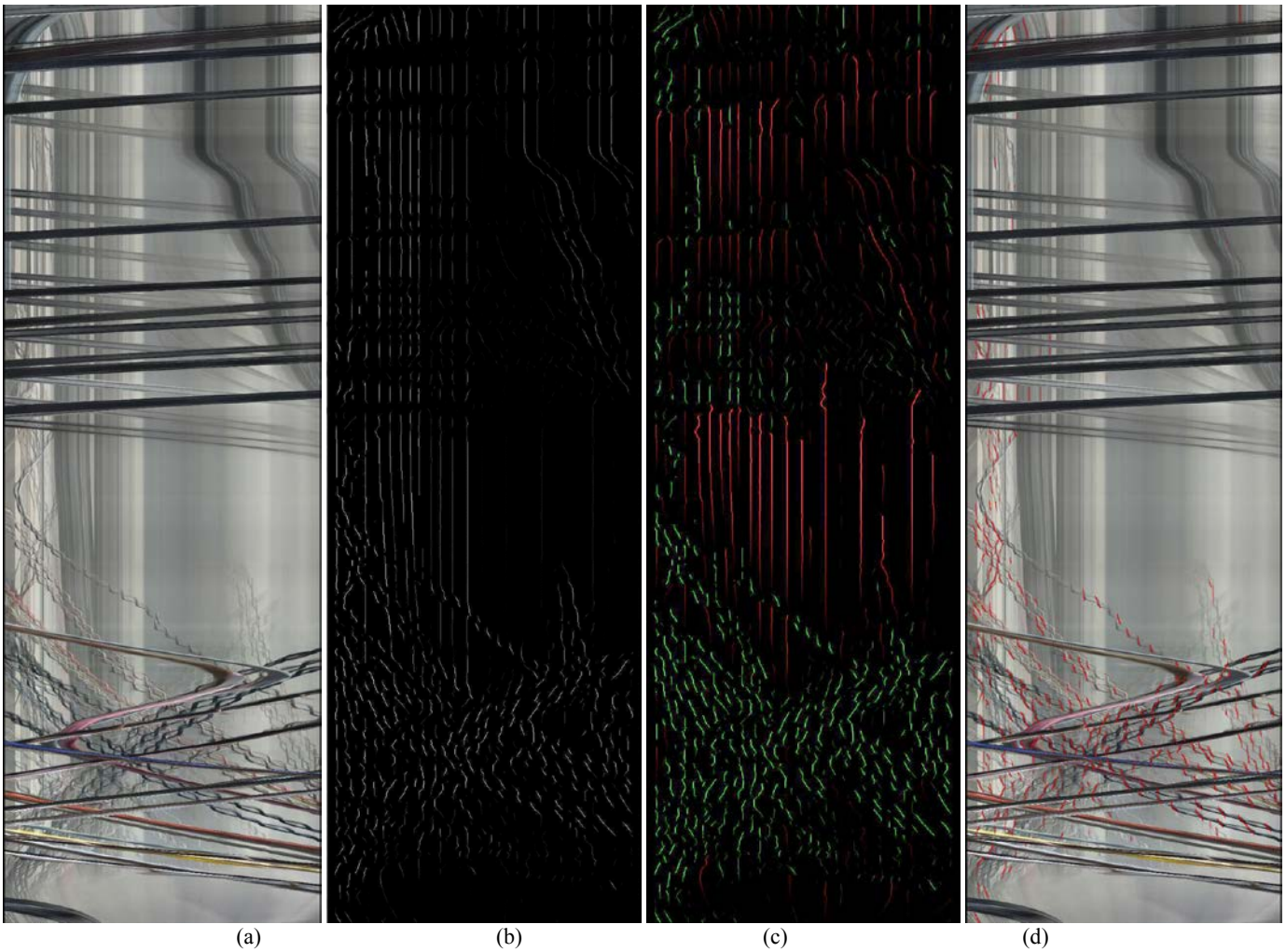
As comparison with previous motion based methods, the HOGTM [17] detects perfect leg crossing on ideal leg-chains where pedestrians passing across a road viewed by a slow

moving camera. No complex situations such as occlusion and heavy trace skewing were dealt with. Although the ideal cases are the most significant for automated driving at local streets, only less than 1/4 of such pedestrians appear in the Caltech dataset; many people are on sidewalks, standing still, or even behind cars. This method compared to [19] results in much robust results as indicated in Fig. 12. The best result yields precision of 81% and sensitivity of 91%. After counting the detected traces and their lengths/frames, ROC curve is also plotted with fewer false positives than [19]. This is because the employed leg stroke has a more global scale than non-smooth corner points used in [19].

Figure 13 displays the detection curves in four zones by using 66 one-minute testing videos (119k frames) in Caltech data set before cross zone action is taken. By changing threshold δ_l over all possible values from minimum 1 to maximum 14, the obtained curves overlay with the those by other shape based methods, where curves drop as the threshold decreases. We can notice that this method is better at close/lower zones than at upper/far zones, because leg parts only have about half size of the entire body and the close one has a sufficient resolution. Overall detection rate is from 50% to 100% as threshold δ_l is lowered down as in Fig. 14. Moreover, the curve shows a lower false positive rate than others even with low δ_l, which makes it better for a fast prescreening method in pedestrian detection. Our algorithm only needs to process four profile images out of a video clip, which makes it more efficient than processing entire images in the whole video volume as other shape based methods.

We found that the method has better results on TASI dataset from 120 cars than Caltech dataset, partially because of the 120 degree wide field of view of HD video, double in resolution, stable outdoor illuminaiton, and lower camera positions for fewer zones to capture motion rather than scanning ground. This is proved in Fig. 13 where close pedestrians have a better detection rate.

The main reasons causing false positives by this filtering approach are: (1) the object stripe frequently occluded by moving objects at closer depths such as a queue of vehicles; (2) a few surface marks on the road. On the other hand, the false negatives mainly come from (a) far distance pedestrians whose leg chain is weak in the motion profile; (b) the trace skew due to fast vehicle turning and translation. Our method has better results in a close depth (lower zones) because of the



(a) (b) (c) (d)

Fig. 11. Detecting leg stripes after tracking trace length in a profile of Fig. 4. (a) Motion profile of one-minute long video, (b) Peak tracking after cascaded filtering, which contains both long and short stripes but all with slow motion. (c) Extracted short stripes in green and ignored long stripes in red. The intensity of two colors show the accumulated stripe length, (d) detected strokes of standing legs in red color overlaid onto motion profile (a).

simple ground surface and large pedestrian legs.

This approach does not require intensive learning, but is more based on physical rules and human learned common walking characteristics. The detection result is less influenced from shape, color, and background in various environments. The disadvantages are (1) its incapablity in detecting humans standing still or whose legs are invisible due to occlusion. (2) detection with a delay for pedestrians to finish actions. The method has to track a leg trace untill it ends, which causes delay at most 30 frames, i.e., one second, designed by our method. The period tracked so far is declared to be a pedestrian at a certain depth. This means an action to avoid pedestrian needs to be taken immediately in real time vehicle control, if the depth is close.

VI. Conclusion

This paper introduced a new motion based method to detect walking pedestrians in driving video. Human leg trajectories are detected and the body traces are confirmed further. We focus on the stopping leg traces in contrast to long and smooth object trajectories, and propagate the information to body traces. The method also works crowd and other cases so that the results are more accurate than previous methods using motion. The implementation is mainly based on the cascade filtering on several motion profile images from video such that a high efficiency for on-board detection system can be achieved. The method has been tested on public video dataset for its effectiveness.

References

[1] P. Reisman, O. Mano, S. Avidan, A. Shashua, "Crowd detection in video sequences", *IEEE-IV* 2004.

[2] P. Dollár, C. Wojek, B. Schiele, P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. PAMI,* 34, 743-761, 2012

[3] M. Enzweiler and D. M. Gavrila, Monocular pedestrian detection: Survey and experiments, *IEEE Trans. PAMI*, 2009.

[4] D. Gerónimo, A. Sappa, A. López, D. Ponsa, "Adaptive image sampling and windows classification for on-board pedestrian detection," *Int. Conf. Computer Vision Systems*, 2007.

[5] Z. Lin and L. Davis, "A pose-invariant descriptor for human detection and segmentation," *ECCV 2008,* 423-436, 2008.

[6] C. Wojek, S. Walk, and B. Schiele, Multi-cue onboard pedestrian detection, *CVPR*, 2009.

[7] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey on pedestrian detection for advanced driver assistance systems," *IEEE PAMI*, vol. 32, no. 7, pp. 1239–1258, 2010

[8] C. Wojek and B. Schiele, "A performance evaluation of single and multi-feature people detection," *Pattern Recognition,* 82-91, 2008.

[9] K. Yang, E. Y. Du, J. Pingge, C. Yaobin, R. Sherony, and H. Takahashi, "Automatic categorization-based multi-stage pedestrian detection," *IEEE Int. Conf. ITS* 2012, 451-456

[10] G. Overett, L. Petersson, N. Brewer, L. Andersson, and N. Pettersson, "A new pedestrian dataset for supervised learning," *IEEE Intelligent Vehicles Symposium, 2008*, 373-378.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE CVPR 2005,* 886-893.

[12] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *IJCV,* vol. 38, 15-33, 2000.

[13] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," *ICCV* 2009, 32-39.

[14] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," *IEEE CVPR*, 2008. 1-8.

[15] K. Yang, E. Y. Du, E. J. Delp, J. Pingge, J. Feng, C. Yaobin, R. Sherony, and H. Takahashi, "An extreme learning machine-based pedestrian detection method," *IEEE Intelligent Vehicles Symposium 2013*. 1404-1409.

[16] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *IJCV,* 63, 153-161, 2005.

[17] M. Kilicarslan, J. Y. Zheng, "Detecting walking pedestrian from leg motion in driving video", *IEEE Int. Conf. Intelligent Transportation System*, 2014, 2924-2929.

[18] M. Kilicarslan, J. Y. Zheng, "Visualizing driving video with temporal profile", *IEEE Intelligent Vehicles* 2014, 1263-1269.

[19] M. Kilicarslan, J. Y. Zheng, A. Algarni, "Pedestrian detection from non-smooth motion", *IEEE Intelligent Vehicle 2015,* , 12(2) 583-595, 2011.

[20] Y. Makihara, "Towards robust gait recognition", *2nd Asian Conference on Pattern Recognition*, 18-22, Japan, Nov. 2013

[21] S. Niyogi, E. Adelson, Analyzing gait with spatiotemporal surfaces, *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, 1994, pp. 64–69.

[22] R. Tian, L. Li, K. Yang, S. Chien, Y. Chen, R. Sherony, "Estimation of the vehicle-pedestrian encounter/conflict risk on the road based on TASI 110-car naturalistic driving data collection. *IEEE IV 2014*, 623-629.

[23] B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," *IEEE CVPR 2005*, 878-885.

[24] C. Harris and M. Stephens (1988). "A combined corner and edge detector", *Proceedings of the 4th Alvey Vision Conference*, 147–151.

[25] A. Ess, B. Leibe, K. Schindler, L. V. Gool, Robust multiperson tracking from a mobile platform. IEEE PAMI, 31(10), 1831-1845, 2009.

[26] Z. Lin and L. Davis, "A pose-invariant descriptor for human detection and segmentation," ECCV 2008, 423-436, 2008.

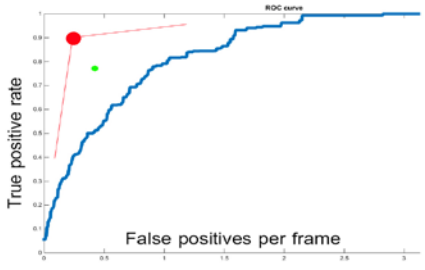


Fig. 12 Comparison with motion based method [17] with ROC in blue. Green dot is the missing rate by our method on the most complex video clip (Fig. 9). Red dot is the AVERAGE missing rates in Caltech video database by our method.

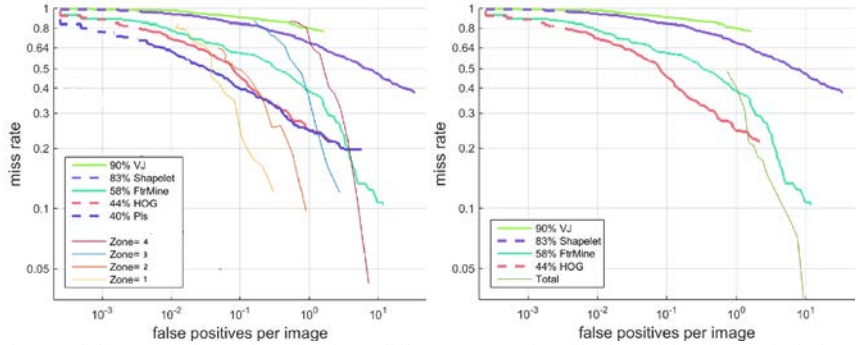


Fig. 13 (left) Comparison of results from different zones after filtering with other typical shape based methods. Fig. 14 (right) A total detection curve of our method from all zones after OR operation of detected pedestrians.