2014 IEEE 17th International Conference on
Intelligent Transportation Systems (ITSC)
October 8-11, 2014. Qingdao, China

Detecting Walking Pedestrians from Leg Motion in Driving Video

M. Kilicarslan, *Student Member, IEEE*, and J. Y. Zheng, *Senior Member, IEEE*

Abstract— **Pedestrian detection in the driving video is an important function for accident avoidance. Different from the detection method based on human shape analysis, this paper introduces a new method to detect walking people from their motion in the driving video. Motion profiles of the driving video are acquired where we found walking people showing their leg moving trajectories as twisted chains. These chains are very different from the moving traces of background and other vehicles appearing as smooth curves according to the vehicle motion mechanism. Thus we design a method to recognize chains uniquely at leg crossing by using HOG features and confirmed with template matching. This method can detect a person in two walking steps. The results show a promising detection rate in the reduced data dimension of video.**

I. Introduction

Pedestrian detection has been an intensively researched topic [1, 2]. Most of them are the appearance based approach that classifies the human shapes in video based on a set of trained features learned from a large number of samples with variations [3, 4]. It is more difficult for a vehicle borne moving camera to detect human because of the dynamic moving background and changeable environment. Pedestrian detection has been studied mainly on detecting the human shapes in individual frames at various resolutions and poses [5, 6]. It has achieved a success while still has a problem of high false positive rate when targets are mixed with complex background. Features such as HOG [7], Haar-like feature [8], LBP [9] are used to describe the characteristics of humans in detection windows. Various classifiers have been tested on comprehensive data set to achieve a balanced accuracy and processing time [10, 11].

Few efforts have been focused on the dynamic motion information in the video for human detection [12]. Gait recognition research works [13] have been investigating walking characteristics of different people in surveillance videos from static cameras for person identification and visualization [14]. The classification mainly uses frequency analysis on the entire sequence. Compared with such static cameras, our work is more challenging because the video is from a vehicle borne camera. Our goal is to detect walking people promptly.

This work tackles the pedestrian detection problem from a new angle. We found that human motion is more random and non-smooth as compared to the background motion generated from the vehicle ego-motion, regardless the complexity of intensity distribution on human appearances in background. This finding allows us to propose a unique method to detect walking pedestrians from their step patterns in a motion profile. By profiling motions from driving video into a temporal image [15], we found that pedestrians' walking action forms chain-type trajectories different from smooth trajectories of other scenes. We therefore developed an original walking step detector to extract the leg moving patterns at their step crossings. The detector applied leg-crossing filtering successfully based on local template matching and HOG features, which yields a high detection rate in the motion profile. The algorithm also works efficiently in a reduced data dimension from video and is suitable to be implemented in real time during the vehicle maneuver in the future. The proposed method can at least improve the accuracy of pedestrian detection if it is combined with the appearance based method. In a good expectation, it can directly search for walking pedestrians, if his/her walks more than one step.

In the following, we start from the image geometry of a vehicle borne camera in Section II, in order to obtain motion profiles at several heights/depths. Then, in Section III, the trajectories between walking pedestrian and background are compared and general leg motions are analyzed in order to model their characteristics. Section IV introduces the methods to detect the leg-crossings and its variations for locating walking trajectories in the motion profile. Section V discusses the experiments and accuracy, as well as efficiency improvement.

II. Collecting Motion Profiles from Driving Video

As shown in Fig. 1, the video frame from a vehicle borne camera is mounted at the height of in-car back mirror, which

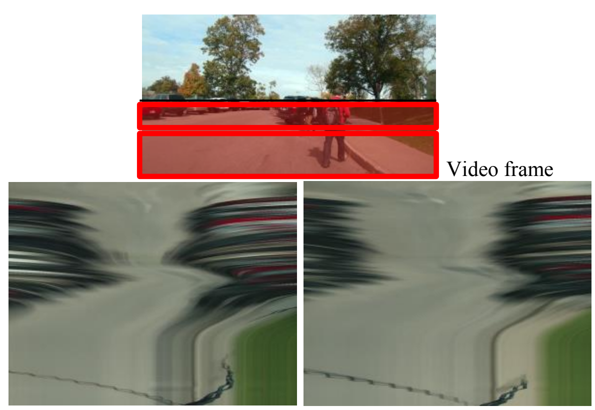


Fig. 1 Video frame and collected motion profiles showing walking trajectories. Two zones in red are selected in the frame above for condensing image intensities into two motion profiles below. The time axis is upward in the profiles and the horizontal axes are the x axes same as in the video frame.

M. Kilicarslan is with the Department of Computer Science, Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202, USA (phone: 812-391-4904; e-mail: mkilicar@iupui.edu).

J. Y. Zheng, is with the Transportation Active Safety Institute (TASI) and the Department of Computer Science, Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202 USA (e-mail: jzheng@cs.iupui.edu).

has a position higher than legs of pedestrians in the 3D space. The projection of the horizon is located in the video frame once the camera is fixed. According to perspective projection rule, the projections of human legs are thus lower than the horizon in the video frame appearing at different heights depending on leg distances/depths. We therefore divide the video frame below the horizon projection into two horizontal zones (shown in Fig. 1) for color value condensing [16]. The heights of zones are determined according to the leg heights in the frames, which are further related to their depths. A distant pedestrian may have legs included in upper zone, but a close pedestrian may only have upper body covered by upper zone; their legs are then captured by the lower zone.

In each zone, pixel values are averaged vertically to produce an array in each frame. The arrays from consecutive frames are connected along the time axis to form a condensed image $I(x,t)$ referred to as *motion profile*, which reflects the spatial and temporal information together in a single image. More detailed process can be found in a recent paper [15]. Through the vertical condensing, many small and non-vertical visual features are blurred out in the motion profile. These features include slanted lane marks in the video frames and other horizontal shapes such as road signs on the ground. Only vertical features such as human bodies and legs, poles, vertical rims of architectures, and side edges of surrounding vehicles are left in the motion profile as motion traces along the time axis (Fig. 2). Even if the observing vehicle has some shaking due to uneven road surface, the instant vertical vibration does not affect the motion traces significantly; it may change the color/intensity and thus contrast in the motion profile in a limited extent (observable in Fig. 2).

In the motion profile $I(x,t)$, we can observe many phenomena including background traces reflected from vehicle ego-motion, as well as the traces of other vehicles, pedestrians, and bicyclists. Many dominant features show their motion in smooth trajectories. If the observing vehicle has a turn in one direction, all the visual features will move in the opposite direction in the video and their traces extend to that direction in the motion profile as well (lower part in Fig. 3a). As the vehicle moves forward, background features expand from the vehicle heading direction in the video frames so that their traces in the motion profile diverge toward sides along the time axis (Fig. 3). For the lower zone to condense the color, it also has an effect to scan the road surface, which means it covers a large scope of the road surface ahead (Fig. 3d). Sometime, it is occupied by a close vehicle in front. The back side of front vehicle may only appear in a certain period and then merge into road surface, which generates blurred shapes dragged along the time axis. The traces walking people appear as chains formed from alternative leg motion in the motion profile. If the camera is set at a relatively lower position and viewing angle is close to horizontal, we can observe the leg motion for a longer time period with fewer zones located in the video frame. The

Fig. 2 A frame (top) from an urban driving video and the motion profile (bottom) vertically condensed from continuous video frames in the red zone (top). The motion profile shows the right turning action of a front vehicle (white) with turning light blinking. Passing vehicles on right are in large curved traces, and pedestrians walking trajectories reveal as chains. The time of each event is recorded accurately in the motion profile along the upward time axis.

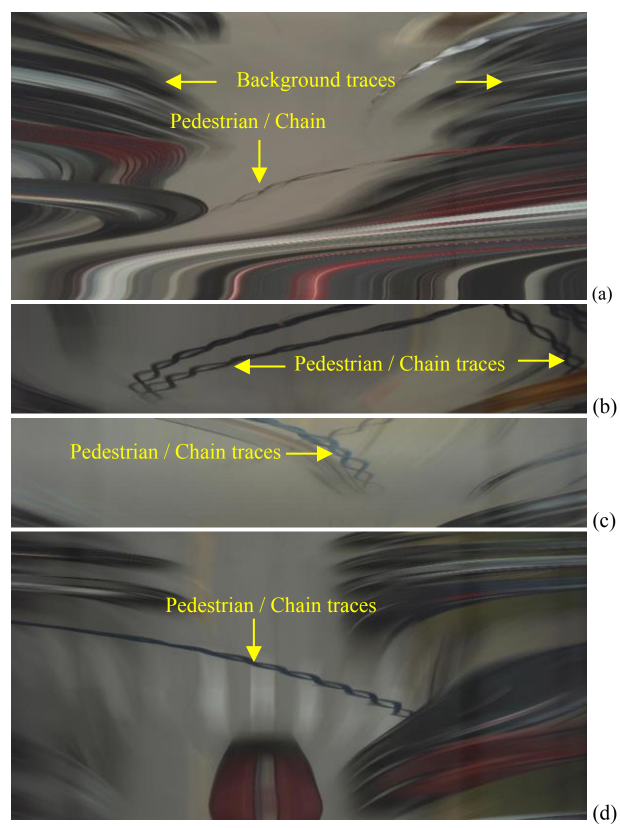


Fig. 3 Various trajectories from walking pedestrians and objects in the motion profiles. The time axis is upward in vertical direction.

goal of this work is then identifying walking traces in the motion profile of video for pedestrian detection and their location understanding during vehicle driving.

III. Motion Trajectories of Walking People

A. Trajectories difference between pedestrians and vehicles

Because a four-wheeled vehicle moves along a smooth path on a flat road, all the scenes relative to the vehicle motion are also smooth when they are projected to the video. Other vehicles appearing in the field of view also have smooth motion, and thus their trajectories in the video are smooth as well. Therefore, we can conclude generally that the trajectories of static scenes and other vehicles are smooth in the motion profile over a long period. This can be observed in the motion profile where local trajectories are highly parallel. Only pedestrian motion produces trajectories that are twisted with varied widths, and walking steps from legs form chain-type traces. This has been confirmed by examining motion profiles in a large driving video database. One example is displayed in Fig. 2. Hence, we set our goal to detect such non-smooth chain trajectories to alarm the existence of walking people in front of vehicle. For human observers, it has no difficulty to distinguish such trajectories from the smooth background traces. Our goal is develop an algorithm to automatically detect parts of such trajectories.

B. Chain modeling from pedestrian motion

In the motion profile, the most prominent character of the walking chain different from other smooth trajectories are (1) leg crossings with different orientations and (2) chain rings with varied step sizes. We design a filter to detect the leg-crossings from their unique structures. We first apply general template matching to find strong trajectories, and then apply HOG operators to detect them in the motion profile. The system will be able to detect a human if he/she walks two steps that producing two leg-crossings, which is different from one crossing of other depth varied scenes such as a pole on roadside against distant background.

Let us model the leg-crossing movement in the motion profile as shown in Fig. 4. Each walking step contains two leg's alternative actions: one has a mild movement and the other steps out largely. Their traces appear as close to vertical and horizontal in the motion profile respectively. These two traces may not have the same color because of the shadow and illumination on legs, but they all have clear contrast from background. A typical template of leg-crossing is synthesized in Fig. 4 according to real samples. The variations of the template are in (a) the vertical scaling along the time axis due to walking and vehicle speeds (video rate fixed), (b) horizontal scaling due to the step size and depth of pedestrians. During the vehicle turning in a certain direction, the scenes are shifted towards the opposite direction in the video such that the template needs to be skewed horizontally in the motion profile. Therefore, we design another template to cover the inversely skewed case where the trace of standing leg moves in the opposite direction to the vehicle turning direction. Figure 5 shows such four detectors (templates) describe the typical walking patterns in one direction corresponding to fast, normal, slow walking speeds, and back skewed traces during vehicle turning. Accordingly, the chain direction (connecting two leg-crossings) can also be specified as 20, 40, 60, and 80 degrees briefly from the *x* axis in the motion profile.

The structure of leg-crossing pattern is scale invariant locally in the motion profile, since the four segments around the crossing point can extend longer in a little larger template window. To locate such leg-crossings, the HOG (histogram of oriented gradient) are computed to generate a sequence of features. The HOG parameters for four templates are divided as in Table I. The block sizes are all in 2×2 cells, and cell sizes are 4×5 pixels. Overlaps between blocks are 50% and 9 bins are prepared in orientations.

Table I HOG Computation Parameters for Templates

Template	Number of blocks
Fast	7×39
Normal	9×29
Slow	9×19
Skewed	9×9

Considering the symmetric case of walking directions, the four templates are further flipped in the *x* direction to create four more templates. With these eight templates, we can detect major leg-crossing spots with different orientations in the motion profile. Figure 6 shows one example of leg-crossing in both direction walking chains. Some leg-crossing spots may be detected by multiple templates. The motion profile is scanned incrementally along the time axis to locate the leg-crossing moments.

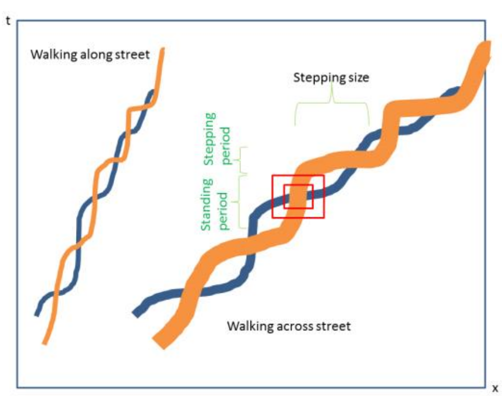
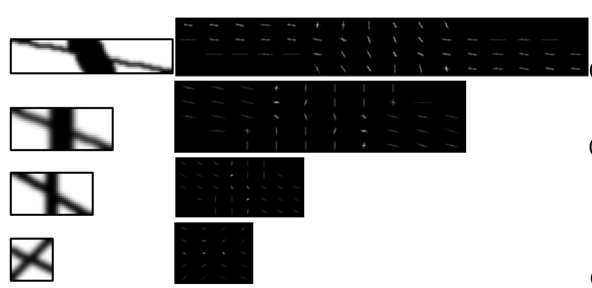


Fig. 4 Typical walking along street and across street are depicted by their trajectories in the motion profile. Two legs stepping alternatively with one closer leg (in orange) occluding the other behind temporally. Red boxes show their scale-invariant local behavior at leg-crossings.



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

Fig. 5 Leg-crossing detectors in the motion profile for the leftward walking chains and their orientation distributions. (a) Fast and close motion temlate has (H×W) 16×80 pixels, (b) Typical crossing with trace pattern, 20×50 pixels, (c) Slow motion template, 20×40 pixels, and (d) Back-skewed template, 20×20 pixels.

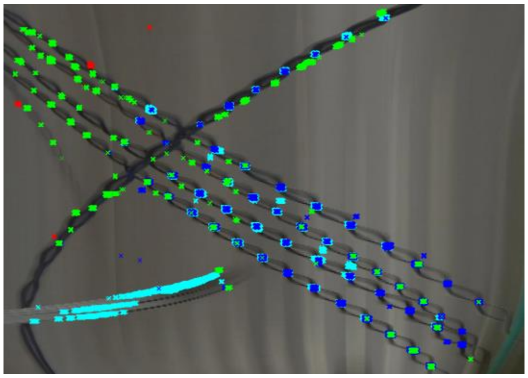


Fig. 6 HOG-only templates detecting chains at leg-crossings in a motion profile. Colors show which template detects the points. Cyan, blue, green, and red indicate the template in 20, 40, 60, or 80 degree orientation detects the leg-crossing, respectively. While HOG locates leg-crossing precisely at various walking speed, it may pick up individual points at some smooth background as well.

IV. Chain Detection at Leg-crossings

A. Intensity template matching for prescreening

Because HOG detects the intensity distribution similar to the template of a structure, it may pick up places with very week gradients as well. Therefore, as a precondition, we further require that the detected chain must have a strong gradient strength consistent to the template. We use the template matching to locate the places with high similarity to the chains. Normalized cross correlation on intensity is applied to the entire motion profile as

$$D(i,j) = \frac{\sum_{i,j}[I(i,j) - \overline{I(i,j)}]\sum_{i,j}[T(i,j) - \overline{T(i,j)}]}{\sqrt{\sum_{i,j}[I(i,j) - \overline{I(i,j)}]^2 \sum_{i,j}[T(i,j) - \overline{T(i,j)}]^2}} \quad (1)$$

where T is template and I is motion profile. The points with the absolute values higher than a threshold are marked, which solves the dark leg traces against bright background, and vice versa. The survived positions are further input into the HOG operator for shape comparison to remove the false positives. Thus, only the positions with similar gradients between the template and motion profile are passed to HOG. This pre-screening also largely reduces the computation time for HOG in the entire motion profile.

B. Leg-crossing detection by HOG features

The HOG features between the template and image spot are compared with their feature distance. A threshold is used to reject the places different from the template. Figure 7 shows the cross-correlation template matching and then the HOG matching result. The prepared eight templates with various orientations are applied to the motion profile individually and their results are combined with Union operation so that false negative rate is reduced.

For a walking chain, at least two leg-crossing spots exist to form a ring. Otherwise, a crossing pattern may also appear at the crossing spot between two traces of static objects. For example, if two narrow objects such as electric poles are at different depths, their motion parallaxes are much different and the traces have different directions in the motion profile. They may be occasionally captured at a crossing spot with one trace over another. However, this type of crossing will never lead any merge again on their traces like a ring on a walking chain. We thus require two leg-crossing spots with one step per leg for confirming the walking action of pedestrian. To examine two leg-crossings that form a ring, the direction of the template is defined as one of 20, 40, 60, and 80 degrees for searching. If we can find another consecutive one along that direction from the first leg-crossing, a ring is confirmed and the human position is identified from the x coordinate in the motion profile.

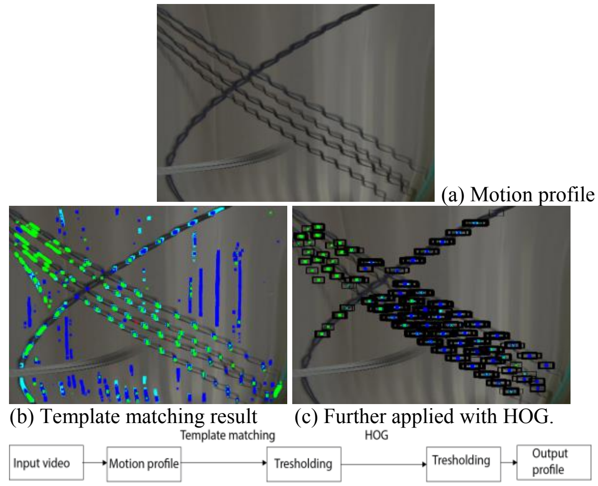


Fig. 7 Prescreening with template matching before HOG computation and HOG confirmation. The black boxes show the template size in the motion profile.

V. Experiments and Discussion

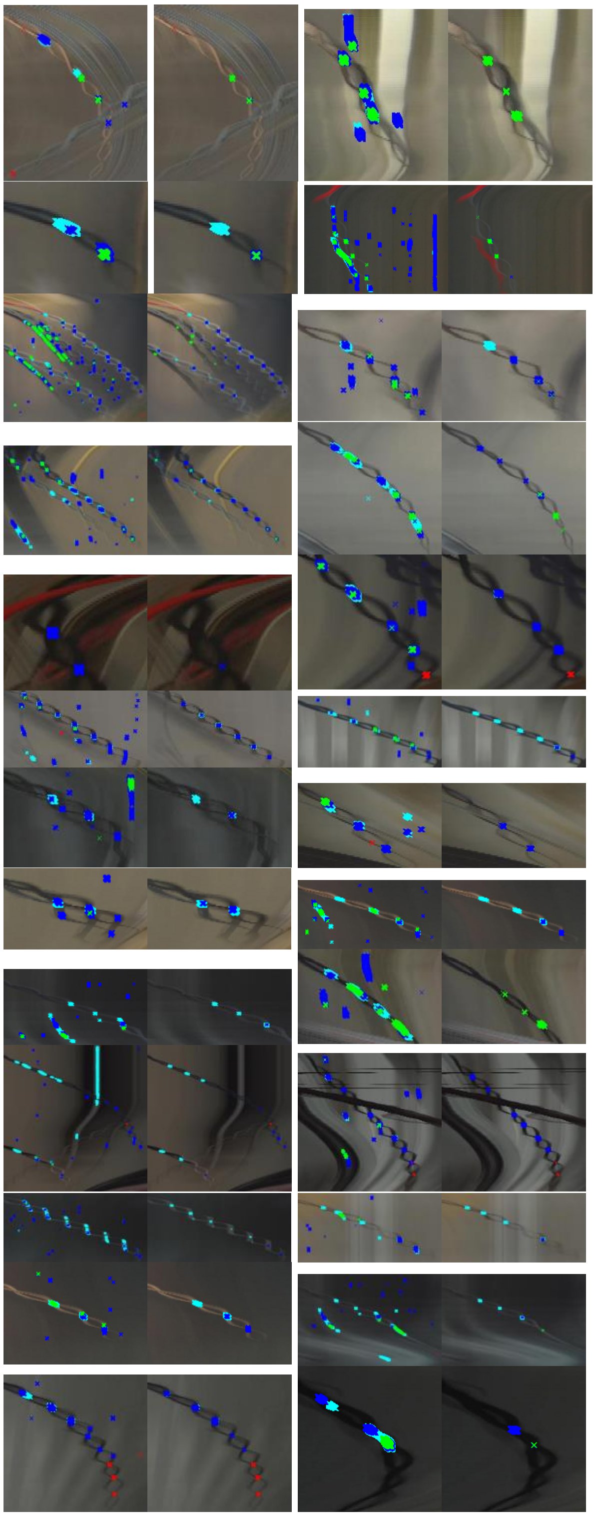
We have set a camera on a vehicle and captured long time driving videos. The motion profile is generated at the rate of 60Hz after the horizon is located in the video frame. Current experiments use two zones for vertical intensity condensing that yields two motion profiles calculated in color. For a leg position or height in the frame imperfectly covered by one zone, it may still generate leg chains with a lower contrast. Even if pedestrian legs are missing in one zone, the second zone should catch the leg movement during walking periods.

The intensity template matching is applied with eight templates for both chain directions and the results are sent to HOG classifiers with the same templates for identification. The first threshold for template matching is set loosely so as to tolerate more candidates, but they do not appear at place with homogeneous intensity distribution or week contrast.

x

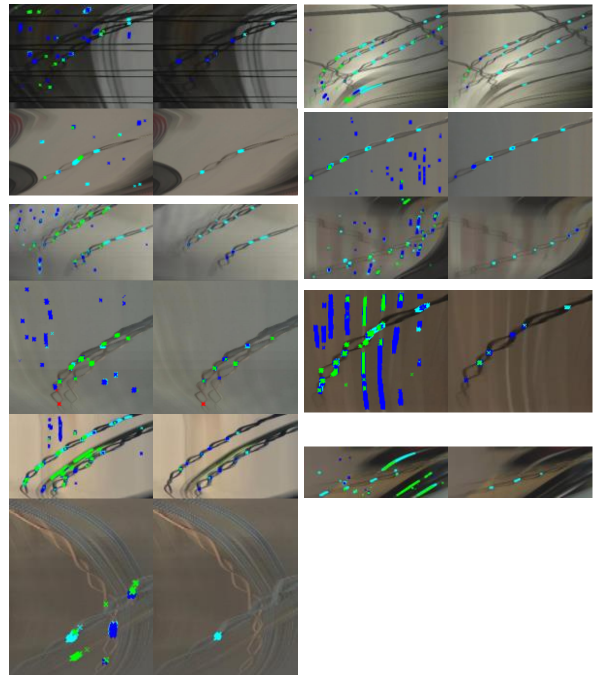
Fig. 8 Successful back chain detection in the motion profile from the lower zone near a street crossing. Cyan and blue are 20 and 40 degree templates respectively. The lower center part is the leaving trace of a front vehicle in red color and white part is from the number plate. After waiting a pedestrian crossed street, the observer vehicle speeded up.

This constrains leg-crossings to be at strong edge places. The detected places are sent to HOG classification module that has a stronger shape and structure discriminating power than the template matching. By applying these steps one after another, we obtain the leg-crossing spots in the motion profile stably. Each spot with several connected pixels is counted as one leg-crossing, even if only one point is



Template, *Then HOG.* *Template,* *Then HOG*

Fig. 9 Samples of leg-crossing detection in leftward walking traces. Gradient matching and then HOG matching results are paired.



Template, *Then HOG*

Fig. 10 The template and HOG detection of leg-crossing spots in rightward walking traces. Spots with a single detected point look dark.

detected. Figure 8 shows a detected chain in the examples.

We have tested 99 motion profiles/clips with pedestrians. We select 87 smaller image patches from them containing clear walking chains. Totally, there are 524 rings (leg-crossings as well) on the chains identified by humans in this set. Others are standing or invisible people due to dark illumination and background. By tuning the thresholds, the algorithms can detect 447 leg-crossings on the trajectories and missing 77 leg-crossings as false negatives. Also, 35 false positives are reported at non-chain spots. The precision and sensitivity are calculated as 92% and 85% respectively. F1 measure is thus 88%. For the entire motion profile larger than these patches, true negative rate may increase significantly but false positives increase at a small rate so that accuracy will become better. The same argument can be derived if we test more and longer videos without pedestrians. Several reasons for obtaining false negative are (1) a crowd of pedestrians where leg traces are mixed such that leg-crossings are interfered by other leg-traces. (2) The leg traces have strong color variation between sun led side and shadow side, which causes a relief-like trace (Fig. 2). Such trace crossing behavior has not been well embedded into our leg-crossing template. A trace tracking mechanism becomes necessary in such a case, since a human observer has no doubt to locate such pedestrian traces.

Figure 9 shows more samples of leftward chain detection. Some points are detected by multi-templates. In the motion profile, background traces are smooth curves during vehicle moving and turning at low speeds. Some horizontal lines in the motion profile are from the motion of windshield wiper in raining days. They have no influences on our detection algorithms in the result. Figure 10 shows similar results on rightward walking trajectories.

Our proposed method cannot detect people standing still aside streets because they have smooth traces as static objects. On the other hand, static behavior at road side has no danger for driving. Most people near driving areas are walking or not absolutely still. Some small body movement may produce traces with different widths, which can be called *non-smooth traces* here. Although our current algorithms based on leg motion have not detected pedestrians walking behind a guiderail or people wearing skirts, those traces are non-smooth curves in the motion profile as well. Further, if the vehicle speed is fast as on a highway, the motion profile is shortened in temporal dimension. A pedestrian walking chain will be squeezed and deformed in the motion profile. This makes the rings invisible or the chain degrade to a non-smooth trace. Most of our testing videos have a low vehicle speed (<40mph) at street junctions and narrow streets where pedestrians appear frequently. Even if the vehicle makes a turn, leg-crossings are detectable by the 80-degree skewed template.

Because we extract the motion profiles from a video clip, the data reduction rate is two lines out of a frame (60Hz sampling), i.e., each frame only has two lines to process for one motion profile. Therefore, this approach tremendously reduces the data size as compared to the appearance based method performed over every frame. In addition, the template matching eliminates many spots for HOG matching so that the processing time is further reduced significantly. This demonstrates that this motion based method has a lower complexity in computation and it may be suitable for real time processing on-board a vehicle. The method can even be combined with appearance based algorithms in the future.

In our experiments, humans can carefully identify pedestrian walking chain in the motion profile in almost 100% accuracy. This is done by tracking leg traces and examining their twisted forms. Our currently designed algorithm focuses on trace crossing for the simplicity and efficiency. This local filtering and classification method avoids tracking and structure analysis of various leg traces.

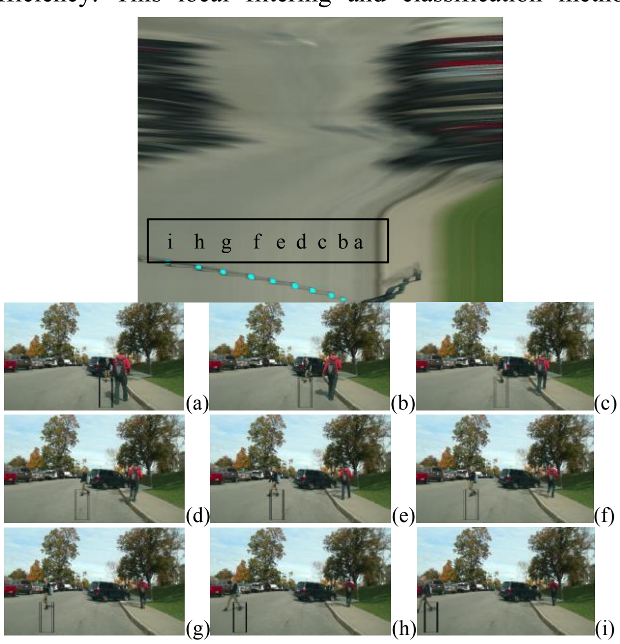


Fig.11 Pedestrian is located in video frames according to the positions of the detected trace in the motion profile condensed lower zone.

VI. Conclusion

By employing the motion profile for driving video, we found that the walking patterns of pedestrian legs as chains are distinguishable from the motion of other scenes in a robust degree. This work thus has a unique focus on identifying pedestrians directly from leg-crossings of walking chains in the motion profile condensed from driving video. We have analyzed the chain type of walking trajectories and designed filter templates to locate leg-crossing patterns in the scene trajectories. This method can extract a pedestrian as long as two steps are detected in the video. The experimental results show a high accuracy in the pedestrian detection as well as the efficiency in using the motion profile.

VII. References

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

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

[3] 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

[4] 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.

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

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

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

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

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

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

[11] 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.

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

[13] Y. Makihara, "Towards robust gait recognition", *2nd ACPR*, pp. 18-22, Naha, Japan, Nov. 2013

[14] S. Bagheri, J.Y. Zheng, "Localized Temporal Profile of Survollance Video", IEEE International Conference on Multimedia and Exposition, 307-313, 2014.

[15] M. Kilicarslan, J. Y. Zheng, "Visualizing driving video with temporal profile", *IEEE Intelligent Vehicles* 2014, 1-7.

[16] A. Jazayeri, H. Cai, J. Y. Zheng, M. Tuceryan, "Vehicle detection and tracking in car video based on motion model", IEEE Trans. ITS, 12(2) 583-595, 2011.