2015 IEEE Intelligent Vehicles Symposium (IV)
June 28 - July 1, 2015. COEX, Seoul, Korea

Pedestrian Detection from Non-smooth Motion

Mehmet Kilicarslan, *Member, IEEE*, Jiang Yu Zheng, *Senior Member, IEEE,* and Aied Algarni

Abstract*— Pedestrian detection has been intensively studied based on appearances for driving safety. Only a few works have explored between-frame optical flow as one of features for human classification. In this paper, however, a new point of view is taken to watch a longer period for non-smooth movement. We explore the pedestrian detection purely based on motion, which is common and intrinsic for all pedestrians regardless of their shape, color, background, etc. We found unique motion characteristics of humans different from rigid objects in motion profiles. Based on the explicit analysis of spatial-temporal behaviors of pedestrians, non-smooth motion points are detected at the motion trajectories of limbs and body. This method works for driving video where both pedestrians and background are moving, and it yields good results as it is less influenced from pedestrian variations in shape and environment. The method also has low computational cost and it can be combined with a shape-based method as pre-screening tool for accuracy and speed.

I. Introduction

Detecting pedestrians in front of a car is critical for avoiding fatal injury. This work tackles pedestrian detection in driving video according to the motion characteristics generated from articulate movements around body. The pedestrian motion is reflected as *chain trajectories* [1] in *motion profiles* [2] condensed from a video. We found that human motion is more random and non-smooth in driving video as compared to background motion generated from the vehicle ego-motion. Such a property is common regardless of varied human appearances. This unique finding allows us to propose new visual cues for detecting pedestrians.

Rigid objects and non-rigid pedestrians have different motion patterns. Because a vehicle borne camera has smooth translation and rotation due to the vehicle mechanism to move along a curve, static objects have smooth motion in the driving video. Their motion trajectories are thus smooth (differentiable), even if the moving direction changes. The change of motion generally occurs in a period at least over one second. Human articulate motion, however, may have a frequency as high as 2Hz in changing directions. An average walking cycle is around 1.26 second, as limbs perform back-and-forth motion around joints. Legs usually do not generate a long slow-motion in order to keep balance. In driving video, limb trajectories are non-smooth around body motion with fast and slow moments alternating. Such movements cause a varied width of pedestrian trace in the time domain.

The contribution of this work is to analyze human walking in the temporal domain and explore pedestrian detection using new visual cues purely based on motion. We first profile motion from video in temporal images called *motion profiles* [2]. Then, we locate pedestrian trajectories different from the smooth trajectories of rigid scenes. Human motion generates non-smooth trajectories due to the self-occlusion by limbs. We identify such moments by examining their trace smoothness. The designed algorithms work efficiently on motion profiles that are reduced from video in data size. This is suitable for real time processing during vehicle maneuver.

II. Related Work

It is more difficult for a vehicle borne camera to detect humans than for a static camera in surveillance because of the moving environment and dynamic illumination. Background subtraction is impossible such that pedestrian detection is explored on the appearance based approach [5, 6, 7, 8, 9, 10, 11]. Such methods classify human by training a set of features using a huge number of samples [12, 13]. Features such as HOG [7], Haar-type feature [15], LBP [16] are used to describe humans shapes in a search window. Various classifiers have also been tested on data sets to achieve a balanced accuracy and processing time [17, 18]. Although such methods have achieved success, they still have problems in high false positive rate and low detection rate when targets are mixed with complex background or under poor illuminations in real scenes [7, 8].

On the other hand, several efforts have been focused on the motion in video for human detection [19]. Between-frame optical flow is combined with shape properties for pedestrian classification. Such noisy flow as the first order derivative is insufficient to characterize walking behavior because entire background and other vehicles are also moving in driving video. A series of gait recognition works [20, 21] investigate walking characteristics of different people in surveillance video for person identification. Compared with static cameras, pedestrian detection is more challenging for vehicle borne cameras.

In contrast to the first order derivatives or image velocity [4, 19], we have found that the motion trace of legs forms a twisted chain in the motion profile [1], and pedestrians are detected if their leg crossings are observed. However, depending on an observing distance, occlusion in crowd [4, 23], variation in walking style, shadows on legs, vehicle/camera motion, and so on, such leg motions may not be captured perfectly. A more general and robust detection method handling body and limbs together is necessary for

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

A. Algarni is with the Department of Computer Science, Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202, USA (e-mail: aalgarni@umail.iu.edu).

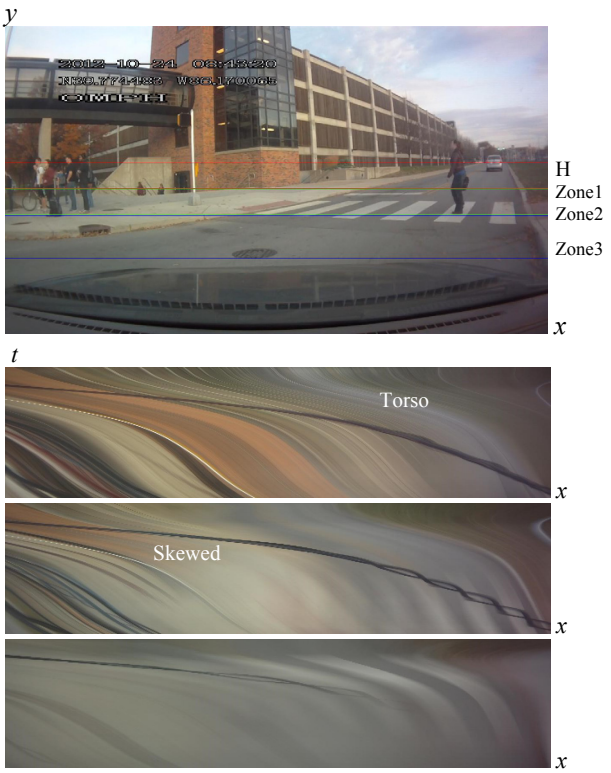


Fig. 1: Five-second motion profiles (lower) obtained from three zones in a driving video of right turn (top). The time axes of profiles are upward. A pedestrian is walking across the street from right to left. Three zones are located below horizon position H in the frame. The pedestrian trace is then skewed severely toward left during the vehicle rotation.

increasing detecting rate. In this paper, our method examines the *Non-smooth Motion* (NSM) of limbs and torso, which corresponds to the second order derivatives and part occlusion.

In the following, we start from the video condensing in Section III to profile motion at several heights covering different depths. In Section IV, pedestrian trajectories are described in their leg and arm-body motions, and NSM point detection is proposed not only on legs, but also on upper body, far pedestrian, and crowds. Section V handles isolated occlusion cases caused by object motion. Section VI is the experiments of our method on accuracy and efficiency.

III. Motion Profiling from Horizontal Zones

A vehicle-borne camera is set at a position near back mirror and is pursuing planar motion on road. Assume the projection of *horizon* is at height H in the video frame. The projections of legs are thus lower than H according to the perspective projection, and some body may appear higher than H depending on its distance. We set several horizontal *zones* at different heights lower than H in the frame for color value condensing as shown in Fig. 1. These zones cover different depths and pedestrian parts including legs, arms and bodies. A lower zone captures the legs of close pedestrians, bodies of very close pedestrians, and ground. An upper zone captures distant humans, bodies of close pedestrians, other moving cars and far background. Pedestrian detection in any zone results in a horizontal position x and a positive count at the frame.

From each zone, a motion profile is condensed vertically from color values to show the long term motion [2]. Pixel values in a zone are averaged vertically to produce an array from each frame. The arrays from consecutive frames are connected along the time axis to form a condensed image, referred to as *motion profile*. The motion profile from zone k, k=1, 2, 3 is obtained as $P_k(x,t) = \Sigma_{y \in zone\_k} I(x,y,t)$.

The motion profiles reflect the spatial-temporal information in a video within several images. Through the vertical condensing, many small and non-vertical features are blurred out in the profile. Such features include slanted lane marks and road signs on the ground. The large scope condensing also smoothens the local motion of natural objects like leaf waving in the resulting profile. Only vertical features such as human bodies and legs, poles, building rims, and vertical edges of surrounding vehicles are left in the motion profile as traces. Even if the observing vehicle shakes on an uneven road, the instant vertical vibration does not affect the motion traces in the profile significantly. Figure 2 shows several examples of motion profiles containing pedestrians at various depths.

The motion profile provides a spatialized way to examine, visualize, and evaluate long term motion that was impossible to estimate by using optical flow. It is particularly suitable for understanding the motion events and behaviors of targets. In profile $P_k(x,t)$ of driving video from an upper zone, background traces reflect the vehicle ego-motion, and the traces of other vehicles are also observable. As a vehicle is turning, all the visual features move in the opposite direction in video and thus their traces also extend to opposite direction in the motion profile. As the vehicle moves forward, background features expand from the heading direction so that their traces in the motion profile diverge toward two sides. A lower zone also scans the road surface. It covers a large scope of road ahead. Occasionally, its motion profile include a front vehicle at rear bumper for a certain period, which generates blurred shapes dragging along the time axis. Nevertheless, these traces are all smooth.

IV. Modeling Non-smooth Trajectories

In contrast to background motion, the traces of walking people appear as chains formed by alternative leg stepping in the motion profile [1]. Although at least one motion profile from a particular zone will cover the leg motion, an upper zone may also cover arms or entire body when a person is far away. For a fast horizontal image velocity generated from vehicle turnning or speeding, the chains are skewed severally to have degenerated crossings and rings. Pedesrians in a crowd [4] have trajectories mixed. Such occlusion yields many intersecting traces with non-smooth points. These have not been modeled in the leg crossing detection by HOG [1].

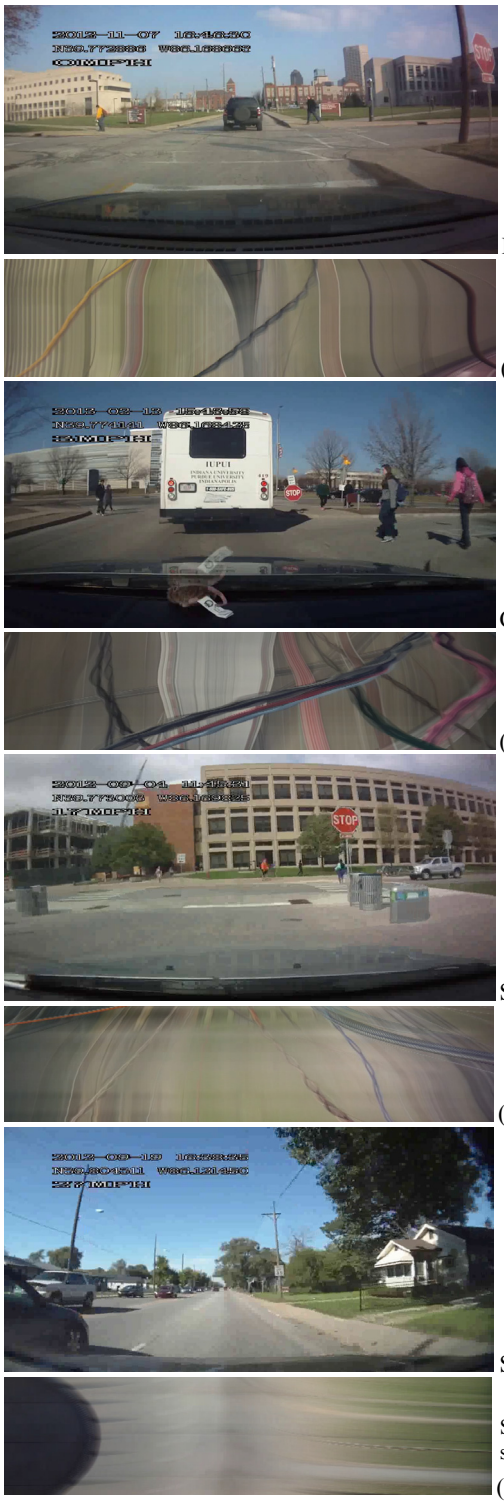
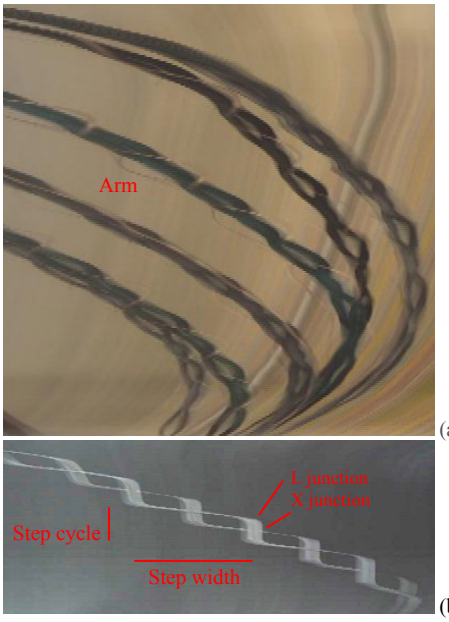


Fig. 2: Video clips of 5 seconds and the motion profiles with the time axes downward. (a)-(c) Upper profiles captures legs as well as bodies, (d) Lower profile catches legs with a fast car speed.

To detect pedestrians more robustly, we further take arms, torso, and their relation into account in addition to legs. Body and arm motion produces twisted trajectories with varied widths in the motion profile. Figure 3 shows leg and arm movements around body; arms show back-and-forth oscillation. The NSM points on traces thus become key for separating pedestrians from rigid objects in background. For pedestrian trajectories as chains in the motion profile, (1) the vertical time span of a step is invariant to the vehicle speed. (2) The horizontal step width is related to the vehicle speed, pedestrian step size, and pedestrian depth in the space.



The time difference between two X junctions is the time for one step (half cycle).

Fig. 3: Pedestrian traces in motion profiles during vehicle motion. (a) Arm traces as thin curves in skin color wave around dark body/leg traces. (b) Typical leg trace walking across street. Two legs step alternatively with X and L junctions. One closer leg occludes the other momentarily.

To mark non-smooth spots on trajectories in profile $P(x,t)$, we compute Hessian matrix [24], which is used in corner detection, from the spatial and temporal derivatives $P_x(x,t)$ and $P_t(x,t)$ in Gaussian window $w(u, v)$ as

$$A(x,t) = \sum_u \sum_v w(u,v) \begin{bmatrix} P_x^2 & P_x P_t \\ P_x P_t & P_t^2 \end{bmatrix} \quad (1)$$

Two eigenvalues $\lambda_1 > \lambda_2$ of the matrix indicate the contrasts orthogonal to local structure and eigenvectors indicate the local structure orientations. We compute the relation of λ_1 and λ_2 to determine whether point (x, t) is a *non-smooth point*, *trace point*, or in homogenous *region* as

$$\begin{array}{rl} \textit{NSM spot / corner} & \lambda_1 \times \lambda_2 > \delta_1 \\ \textit{Smooth trace} & \lambda_1 > \delta_2 \\ \textit{Region} & \lambda_1 < \delta_2 \end{array} \quad (2)$$

where $\delta_1 > \delta_2$ are thresholds. We further mark peaks at local maximum of $\lambda_1 \times \lambda_2$ value for counting the number of corners. The NSM detection is more powerful than leg crossing detection [1] in finding irregular and varied pedestrian traces.

Figure 4 shows the detection of NSM on leg crossings in motion profiles. Leg crossings on pedestrian chains are yielded from the occlusion between two legs. They are identified by HOG in [1] and are also detectable here by NSM algorithm as X and L junctions. L junctions are at stop or start moments of a limb action/stroke in each step. This

means that the NSM extraction is more general than the HOG detection of leg crossings. Further, T junctions are detected by the NSM method not only at mixed traces from crowd and the occlusion of arms over torso, but also at the occlusion between rigid objects. Figure 5 shows additional NSM points easily missed as regular leg crossings.

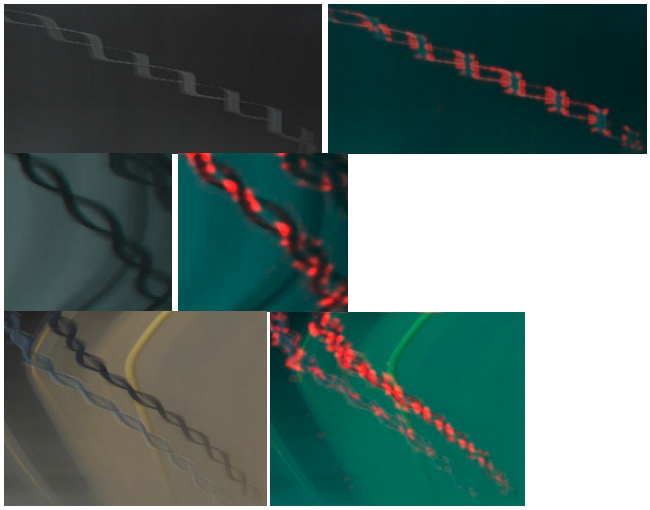


Fig. 4: NSM spots detected in red at crossings on leg chains.

V. Pedestrian Detection and Tracking

A driving video contains the motion of both background and dynamic vehicles. Although their traces are all smooth in the motion profiles, their occlusions in the space cause trace intersections as T junctions in the motion profile. Occasionally, static objects with a large depth variation produce Y junctions in the motion profile from an upper zone. Although the probability of detecting such occlusion spots is not as high as finding NSM moments on pedestrians, these events on object boundaries are false-positives (FP) in pedestrian detection.

We track traces in the motion profile to remove those isolated FP points. Figure 6 shows NSM points on a pedestrian trace (red) against background traces (yellow) detected by using (2). The difference between pedestrians and isolated object occlusions exists in the density of NSM points. The NSM points on pedestrian traces are more random, denser, and wider in distribution, while the vehicle occlusions as T junctions are sparse and isolated on smooth traces. Moreover, T junctions from occlusion will not lead further trace merging as the ring on a pedestrian chain, i.e., an NSM point from occlusion separates two trace branches completely. We can observe that the NSM detection obtains accurate peaks because of the nice local behavior of Harris corner detector. We thus track the traces between NSM points. Both local structure and tracked history are used. Our approach starts tracking after a point with large λ_I is discovered in the motion profile. We employ Hidden Markov Model (HMM) to describe the probabilities on trace as either smoothly moving object/vehicle or pedestrian (the two states defined in HMM). The transition between states are set based on NSM extraction, i.e., meeting NSM points around 0.5~0.7 second on trace adds more probability to the current state as pedestrian, while tracking on a smooth trace increases the probability as rigid object. The sparse NSM distribution on an object trace reduces the probability as pedestrian quickly, while dense NSM points keep the state of trace as pedestrian.

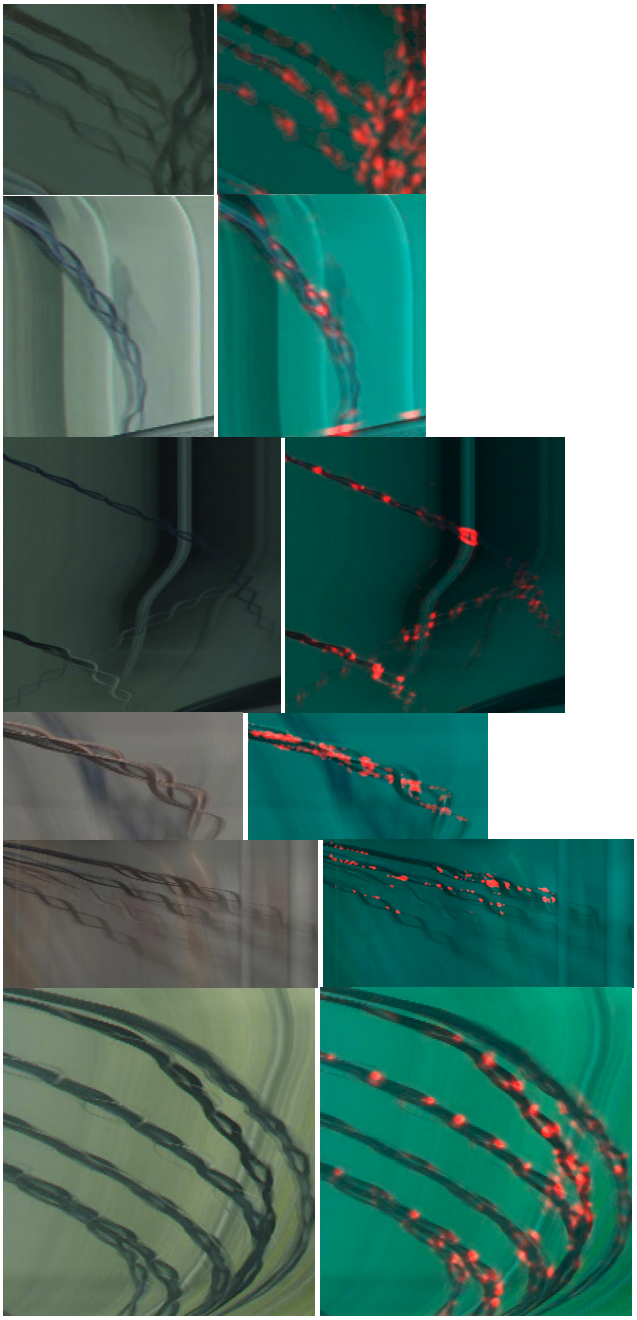
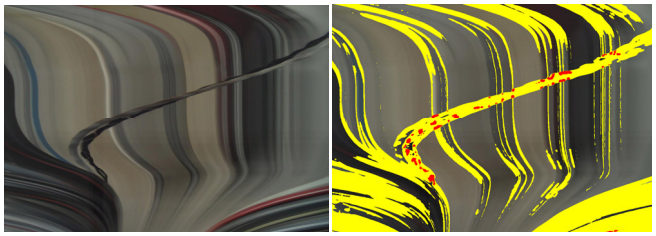


Fig. 5: Segments of motion profiles (left) and detected NSM spots (right). Traces are not regular leg crossings on pedestrian chains. Smooth traces of rigid objects in background are excluded from NSM detection.

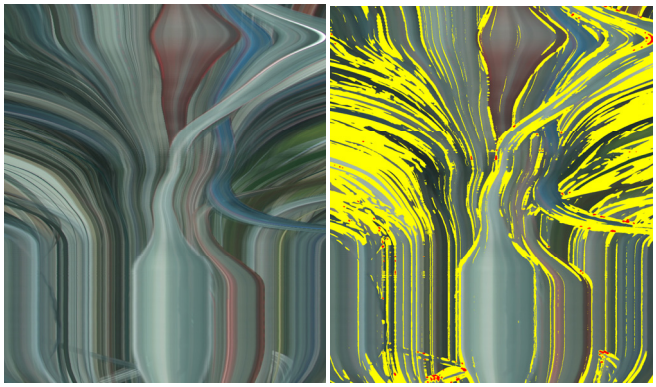
VI. Experiments and Discussion

In our experiments, we have used the pedestrian database of TASI 110 Car Naturalistic Driving Video [22]. AVI videos with size 1280×720 pixels contain fast and slow motion as well as stop period of vehicles. The motion

profiles are generated at the rate of 30Hz from these videos. We select 50 typical clips (each is 5 second long) containing pedestrians who are crossing street, along street on sidewalk, in parking area, during vehicle turning, at very close depth (<5m), night scene, etc. Two zones, which has 50 pixels for upper zone and 100 pixels for lower zone, are set to cover close scenes around 5~10m and distant scenes beyond 10m, respectively. A larger zone in height can blur traces more and remove noises, but also reduces the contrast of pedestrian traces, which has minor effect to NSM detection, since low contrast corners to a certain degree are still detectable on the traces. Even if the leg position or height is not perfectly covered by the lower zone, the motion profiling from the upper zone may still catch leg chains partially and upper body.



(a) Motion profile NSM points and traces



(b) Motion profile NSM points and traces

Fig. 6: Detected NSMs and smooth traces in the profiles from an upper zone. (a) A vehicle/camera slowed down at street corner to wait pedestrian passing and then turned rightward. (b) Many vehicles in front of the camera left traces during stopping period and then drove away on a busy street.

In general, it has no problem for a human observer to distinguish pedestrians from smoothly moving background in a motion profile. The cues used in this separation by humans are non-smooth trajectories and chain trajectories of pedestrians. Counting the peaks of $\lambda_1 \times \lambda_2$, we evaluate the accuracy of NSM algorithm from the motion profiles. The NSM method detects more pedestrian traces than leg crossings by HOG [1]. Up to 98% of the corners on pedestrian traces are marked as NSM peaks; the sensitivity of the method is much higher than the leg crossing detection by HOG. The method has better results in detecting crossing-street pedestrians than in finding pedestrians on sidewalk from fast moving vehicles. Close pedestrians are more accurately detected than pedestrians far away for their large sizes of chain traces and less influences from far objects. The false negatives (FN) are from (1) far pedestrians whose NSM of bodies is invisible in the profile, or small legs yield weak contrast in upper profile, (2) sidewalk pedestrians with traces skewed severely due to a fast vehicle turning/motion.

The false positives (FP) arise not only from (a) between-object occlusion in upper zone at crowded streets, but also from (b) temporal alias of fine patterns such as fence, (c) stair-type zigzags on trace edges from fast driving or turning vehicle/camera, due to a low temporal sampling rate (30Hz). Some heavy video compression, e.g., MPEG, generates false positives as well. Therefore, AVI format is used in our experiments. The motion profile from a lower zone may occasionally pick up (d) ground surface marks if the threshold δ_I is set too low. Instantaneous light changes such as (e) other car's light blinking, (f) highlight reflection and headlight casted from other cars, and (g) going through strong shadows also generate NSM points as false positives. Window wiping in raining days causes NSM points in the profile. Most false positives appear as T junctions on the traces. The leg crossing detection [1] is stricter to prevent the inclusion of these side phenomena. Overall, the NSM detection covers more irregular pedestrian traces, but also include more false positives. Although the accuracy of NSM method is scene dependent, we obtain the ROC curve by counting corner peaks along the pedestrian traces from the 50 profiles (Fig. 7). Associated windows to the detected peaks are 15 pixel wide against the rejected pixels in the profile (640×150 pixels). Also, Figure 8 shows some phenomena in complex profiles.

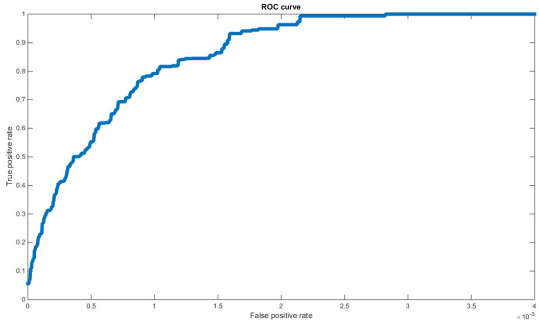


Fig. 7: ROC curve of pedestrian detection from NSM points. The horizontal coordinate displays FPR ranging in [0, 4×10^{-3}).

In our experiments, we found that 30Hz sampling rate is insufficient for the scenes with high image velocity. This results in *aliasing* on the traces and *zigzags* on slanted or close-to-horizontal traces in the motion profile. To prevent these defects, median filtering or blurring has been applied to the motion profile temporally to remove the artifacts. The pedestrian trajectories are then less affected from the zigzags because of their larger scale than the digitized errors.

Our method cannot detect any stand-still human since no motion is revealed; a shape based method [18] works in such situations. On the other hand, the proposed method will not be fooled by any painted human on advertisements and is less confused with similar shapes to pedestrians in the scenes. It is stronger than a shape based method in detecting

pedestrians wearing irregular costumes like long skirts and large hats, carrying large bags, and partially occluded pedestrians, because they may still generate non-smooth trajectories.

Temporally, our NSM based methods has a delay to report pedestrians, which is several frames corresponding to the half size of corner detecting window. Under 30Hz sampling of video, the delay is 6 frames (0.2 second) for NSM method due to derivative computation in Gaussian window. Because we extract the motion profiles from a video clip, the data reduction rate is several lines out of a frame. This approach tremendously reduces the data size as compared to the appearance based method in principle.

VII. Conclusion

By condensing videos to the motion profiles, we found that the walking patterns of pedestrians are distinguishable directly as non-smooth traces from the motion of other scenes. This work has a unique focus on identifying pedestrians robustly from non-smooth traces in the motion profiles. We analyze the walking trajectories and extract non-smooth points on the scene trajectories. Unlike statistical learning methods, we explored underlying physical principles and developed methods to extract such new visual cues. The method has less cost in classification than appearance based approaches, and is efficient by using the motion profiles. It has a higher sensitivity than the HOG based method for detecting leg crossings.

References

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

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

[3] 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.

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

[5] 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

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

[7] 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.

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

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

[10] 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.

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

[12] 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

[13] 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.

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

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

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

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

[18] 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.

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

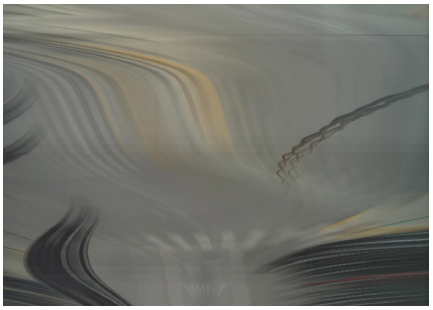
[20] Y. Makihara, "Towards robust gait recognition", *2nd Asian Conference on Pattern Recognition*, 18-22, Japan, Nov. 2013

[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.

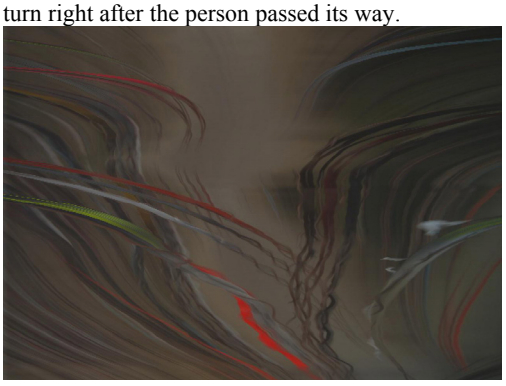
[22] R. Tian, L. Li, K. Yang, S. Chien, Y. Chen, R. Sherony, "Estimation of the vehicle-pedestrian encounter/conflict risk on the road based on TASI 110-car naturalistic driving data collection. *IEEE IV 2014*, 623-629.

[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.



(a) A lower profile showing a person approaching so that his legs are profiled first. The car/camera slows down at middle and then turn right after the person passed its way.



(b) Multiple pedestrians walking in front of a slow vehicle. Their directions are along the street. Body and legs are visible in profile.

Fig. 8 Examples of complex motion profiles for 15 seconds.