Early Detection of the Pedestrian's Intention to Cross the Street

Sebastian Köhler, Michael Goldhammer, Sebastian Bauer, Konrad Doll, Ulrich Brunsmann, Klaus Dietmayer

Abstract— **This paper focuses on monocular-video-based stationary detection of the pedestrian's intention to enter the traffic lane. We propose a motion contour image based HOG-like descriptor, MCHOG, and a machine learning algorithm that reaches the decision at an accuracy of 99 % within the initial step at the curb of smart infrastructure. MCHOG implicitly comprises the body language of gait initiation, especially the body bending and the spread of legs. In a case study at laboratory conditions we present ROC performance data and an evaluation of the span of time necessary for recognition. While MCHOG in special cases indicates detection of the intention before the whole body moves, on average it allows for detection of the movement within 6 frames at a frame rate of 50 Hz and an accuracy of 80 %. Feasibility of the method in a real world intersection scenario is demonstrated.**

I. Introduction

Early detection of pedestrians entering a traffic lane in an urban traffic environment is a current challenge in order to increase vehicle safety. The worldwide traffic crash statistics published by the World Health Organization in 2009 reports that half of 1.27 million victims are vulnerable road users (pedestrians, motorcyclists and cyclists) [1]. Even in Europe and the U.S., which rank first in road infrastructure safety and where fatal injuries decrease continuously, the number of pedestrians involved in severe injuries is still high. Hence, pedestrian detection is becoming an integral part of advanced driver assistance systems (ADAS). Due to the ability of a pedestrian at the sidewalk to suddenly start a motion or to change the direction of motion towards the lane, a dangerous situation may occur within some hundreds of milliseconds. Therefore, an ADAS should not only issue a warning to the driver but also initiate autonomous braking or maneuvering for collision avoidance, when the driver is no longer able to react in time. This requires early and reliable detection of dangerous pedestrian movements which may be recognized either vehicle-based or infrastructure-based.

Many dangerous situations due to occlusions, that cannot be detected from a vehicle, occur at intersections. Therefore, several research projects worldwide, e.g. the American IntelliDrive program [2], the European SAFESPOT [3] and INTERSAFE [4] projects or the German Ko-PER project of the Ko-FAS research initiative [5], address infrastructure-based pedestrian perception aiming at an improvement of road safety by combining infrastructure information with local vehicle data. Video-sensors are commonly used for high-speed and high-resolution data acquisition and pedestrian recognition in urban traffic scenarios using machine learning algorithms is in the focus of research since some decades. Nevertheless, little is known about early indicators which may lead to a quick decision, if a pedestrian at the curb starts to enter the lane or not. Note, that already a gain of a few hundred milliseconds may prevent severe injuries [6]. The main contribution of this paper is to propose an appropriate video-based descriptor extracted from the grayscale image of a stationary monocular camera and a machine learning technique for reaching the decision.

II. Related Work

Video-based observation and interpretation of pedestrian movement requires in a first step human detection. State-of-the-art ADAS detectors often combine two complementary sensors in order to provide features, range and region of interest (ROI), as e.g. provided by stereo-video [7], [8], [9], or by data fusion of a lidar sensor and a monocular camera [10], [11]. Reviews of established video-based pedestrian detectors comprising descriptors, classifiers and benchmarks are presented in [12], [13], [14]; let us further refer to [15] for a survey of state-of-the-art pedestrian detection. The extraction of pedestrian descriptors for a complete frame, the stereo vision and classification algorithms typically are computationally intensive. Hence, some authors have presented real-time detection systems using dedicated algorithms on the CPU [16], or choosing FPGA or GPU hardware for acceleration [17], [18], [19]. Multi-sensor systems for real-time intersection monitoring have been proposed e.g. in [11], [20], [21].

The parameters of gait are well known from human biomechanics research. It has early been shown, that the acceleration within the first stride already results in the mean velocity of normal walking [22], which lies in the scale of 1.4 m/s at intersection crosswalks [23]. The mean velocity is proportional to the product of stride length and stride frequency, for which medical gait analysis studies have reported typical values of 1.2 m and 0.9 Hz, respectively. For an analysis of basic gait parameters see [24], [25]. Vision-based human motion analysis is reviewed e.g. by [26], [27]. Especially, for capture and classification of gait, motion history images (MHIs) [28] of stride [29], gait [30] and frame difference [31] have been proposed that emphasize

S. Köhler, M. Goldhammer, K. Doll, U. Brunsmann are with Faculty of Engineering, University of Applied Sciences Aschaffenburg, Aschaffenburg, Germany sebastian.koehler@h-ab.de, michael.goldhammer@h-ab.de, konrad.doll@h-ab.de, ulrich.brunsmann@h-ab.de

Sebastian Bauer is with Pattern Recognition Lab, Departement of Computer Science, University Erlangen-Nuremberg, Erlangen, Germany sebastian.bauer@informatik.uni-erlangen.de

K. Dietmayer is with the Institute of Measurement, Control, and Microtechnology, Ulm University, Ulm, Germany klaus.dietmayer@uni-ulm.de

the weight of successive contours in gait recognition. For a comprehensive review of these works, see [32].

In contrast to the application of these methods to gait recognition and analysis, there is still few research on the action intentions of pedestrians that aims at the development of ADAS. Early approaches using Kalman filters (KF) are trajectory-based, as e.g. in [33], including interacting multiple model filters in order to account for the ability of humans to suddenly change their type of motion [34]. Schmidt and Färber, however, presented experimental studies, from which they conclude that for a human observer parameters of body language such as leg or head movements are indispensable for the decision, if a pedestrian at the curb enters the lane [35], [36]. They found that the sole consideration of trajectories is insufficient. Though a technical system may overcome the limits of human perception with respect to spatial and temporal resolution, the study demonstrates that there are early indicators beyond the physical parameters of a trajectory. In [37] some indicators are proposed. Recently, Keller, Hermes and Gavrila presented a probabilistic path prediction for pedestrians walking towards the road curbside [6]. Using learned motion features gathered from the dense optical flow of a car-based stereo-system they obtained a classification performance for walking vs. stopping better than that of state-of-the-art KF systems. In [38] a complete active pedestrian safety system is introduced.

In this paper we focus on scenarios, where a pedestrian stands at the pavement and decides to enter the lane. As his or her acceleration is generated predominantly by the legs, we expect, motivated by the results of biomechanics, that the essential information of the pedestrians movement within the first stride is encoded in the legs' movement within the first step. However, bending of the upper trunk forward may indicate the start of walking some hundred milliseconds earlier [39] and a pedestrian at the curb may not stand still before, even if a traffic light shows red. The younger the person, the more dynamics might be expected. Thus, we take into account the language of the whole body before and during the first steps. On this basis, we propose a HOG-like monocular-video-based descriptor in combination with support vector machine (SVM) classification, that allows to decide within the time slot of the first step of the initial stride, if a person starts walking. Our algorithm is solely based on the body poses of the initial movement which are encoded in motion contour based features. The algorithm is designed for stationary intersection monitoring. In a case study at laboratory conditions we evaluate its performance with respect to classification and response time. We demonstrate feasibility of the method in a real world intersection scenario.

The paper is organized as follows: in Section III we describe the processing chain of the proposed method and we present our implementation. Both performance characteristics evaluated on a case study and feasibility in a real world scenario are presented in Section IV before we summarize the main conclusions and discuss open issues in Section V.

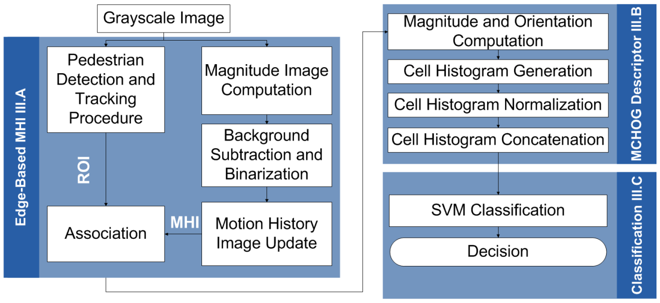

Grayscale Image
Edge-Based MHI III.A
Pedestrian Detection and Tracking Procedure
Magnitude Image Computation
Background Subtraction and Binarization
ROI
Association
MHI
Motion History Image Update
Magnitude and Orientation Computation
Cell Histogram Generation
Cell Histogram Normalization
Cell Histogram Concatenation
MCHOG Descriptor III.B
SVM Classification
Decision
Classification III.C

Fig. 1. Methodology of motion classification via MCHOG; labels III.A - III.C refer to the following sections.

III. Method

It is expected that the subregion of the image that covers the pedestrian within its bounding box is available for a time series, e.g. by the fusion of LIDAR- and video-data and a HOG-based detection [11]. The methodology of our approach to generate the motion descriptors within this box and to classify the motion is illustrated in Fig. 1. The approach is composed of three major stages. In the first stage, which operates on grayscale images, an edge-image is obtained by directional differentiation and an MHI is composed of consecutive edge-images. Building on this MHI representation, in the descriptor (MCHOG) stage normalized cell histograms of oriented gradients are calculated, which are eventually concatenated and form the feature descriptor. In the final stage the descriptors are classified using a linear SVM.

We start with an outline before we go into details. Our MCHOG is based on local directional changes of contour patches extracted from MHIs. Since robust closed contour extraction may become a challenge in cluttered urban traffic scenarios, we analyze patches which are evenly distributed over the pedestrian ROI, making the descriptor less sensitive to gaps in the contour. At first we compute an edge-image by differentiation yielding gradient magnitudes and orientations. A static background edge image is used for subtraction in order to generate a foreground image. With our indoor case study the background can be captured in advance. For our outdoor scenes, however, the background must be updated recursively during run-time. We have implemented the mixture of Gaussians (MOG) model [40] for that purpose on the GPU including a grayscale shadow elimination [41].

The usage of a shadow elimination algorithm, however, proved not to be necessary for the indoor case study, because the magnitude of shadow edges is rather small and these edges are suppressed by thresholding for binarization. The resulting binarized edge image is propagated to the update pipeline of the MHI. Afterwards, the MCHOG descriptor is computed.

A. Edge-Based MHI

The magnitude $|I_{edge}(x,y)|$ of the gradient images I_{edgeX} and I_{edgeY} at each pixel (x,y) is generated by convolving the grayscale input image I with the differential kernels $K_x = \begin{pmatrix}-1 & 0 & 1\end{pmatrix}$ for the x-direction and $K_y = K_x^T$ for the y-direction

$$I_{edgeX} = I_{gray} * K_x \tag{1}$$

$$I_{edgeY} = I_{gray} * K_y \tag{2}$$

and by applying the L2-norm. The next step is background subtraction from the gradient magnitude image. Binarization with a small threshold eliminates noise and smooth shadow. Let $\Psi(I(x,y,t))$ be the respective binarization of an image sequence $I(x,y,t)$, t the frame number and τ a decay value. The MHI-intensity $H_\tau(x,y,t)$ is computed using the update rule

$$H_\tau(x,y,t) = \begin{cases} \tau & \text{if } \Psi(I(x,y,t)) \neq 0 \\ max(0, H_\tau(x,y,t-1) - 1) & \text{otherwise} \end{cases} \tag{3}$$

The pixel intensities in the MHI are set to zero if they are older than the decay value τ. This update function is called for every new video frame analyzed in the sequence. Depending on the value chosen for the decay parameter τ, an MHI can encode a wide history of movement [28], [32]. We have empirically found that $\tau = 10$ represents an adequate trade-off for our application.

B. MCHOG Descriptor

When a pedestrian intends to walk, many parts of the person move accordingly. To capture these different local motions we propose to use a HOG descriptor of the edge-based MHI with appropriate adaptions to the original implementation [42]. As Dalal and Triggs we compute the magnitude and orientation of the gradients, divide the detection window into cells and compute cell histograms [42]. In contrast to the original HOG descriptor we do not perform a normalization of blocks of cells which is used conventionally to boost invariance against illumination changes and foreground-background contrast. In our case a block normalization scheme affects intention recognition rather adversely, because it reduces the local difference between neighboring cells whereas the very same should be captured. Instead, we only normalize the cell histograms by applying an L2-Hys-norm. Finally, concatenating all normalized cell histograms yields the MCHOG descriptor. The default descriptor is optimized for a detection window of 384 × 704 px corresponding to the average bounding box of pedestrians in our database. We use a cell size of 32 × 32 px, 12 bins for cell histogram quantization and a hysteresis threshold of 0.2 yielding a 3168-dimensional MCHOG feature descriptor.

C. Classification

We employ a linear 2-class SVM for classification purposes. A parameter search in order to find a suitable penalty multiplier C was performed using an exponentially growing sequence $C = \{2^{-10}, 2^{-9.5}, \ldots, 2^{15}\}$.

IV. Experiments and Results

A. Data Acquisition

We generated a video database comprising 170 videos of 26 adult test persons, male and female, standing upright and start walking at some point of time. This database was used for the development of the descriptors and finding the optimal descriptor parameter set for machine learning and for evaluation. The videos were captured at daylight in a laboratory environment using a CMOS high-speed camera, operated with a resolution of 1128 × 752 px at a frame rate of 50 fps. This allows us to resolve 10 - 20 single shots within the first step. A complete sequence covers a period of about 15 s. The laboratory features a huge glass front and roof, enabling the lighting to be as natural as possible in an indoor environment. We have not applied any artificial lighting, the color and texture of the carpet fairly resembles tarmac.

B. Scenarios

We specify a set of five scenarios with typical motion activities. To support the orientation of the test person, we attached tape markers to the floor representing position 1-5 and the approximated location of the curb between road (on the right side) and pavement (left), see Fig. 2.

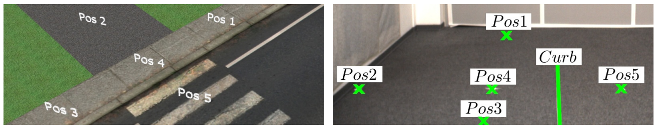


Fig. 2. Test field: laboratory floor (right picture) and the respective curb (left picture).

The entire width of the test corridor is about 4 m, 1.5 m representing the pavement (left) and the remaining 2.5 m representing the road (right). The positions 1-3 mark the initial positions of scenario $s_1 - s_3$, each about 1.5 m away from the center position 4. Both locations 4 and 5 play different roles in each scenario. The particular action sequence of the scenarios is described below.

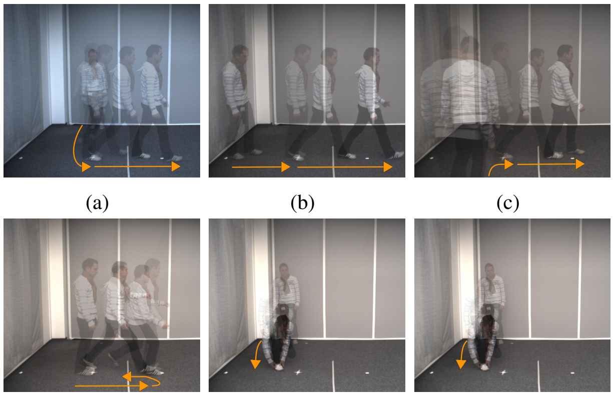
1) Scenario $s_1 - s_3$ (Fig. 3 a-c): The test person walks from the starting position 1 (2, 3) to the center (position 4) and turns in direction to the road. This reorientation corresponds to the fact, that a person willing to cross the road commonly observes the state of the traffic. About two seconds later the pedestrian decides to cross the road and moves into direction of position 5.

2) Scenario s_4 (Fig. 3 d): A pedestrian is starting to cross the road, but suddenly he or she recognizes a vehicle approaching. As a reflex, the person reverses the direction of movement in order to reach the pavement again. A few seconds later the pedestrian crosses the road without disruption.

3) Scenario s_5 (Fig. 3 e-f): For this scenario, the subject was not given any particular instruction. The test person can do any kind of action while waiting for a good opportunity to cross the street. He or she can do some stretching (pretending to be a jogger), place a mobile phone call, walk up and down the pavement, tie his or her shoes, et cetera. At an arbitrary point of time, the test person crosses the road.

C. MCHOG/SVM Classification Results

The evaluation was carried out on data of the test persons performing the 5 scenarios. 4663 pedestrian ROIs were automatically extracted using a multi-scale HOG descriptor



(a) (b) (c)

(d) (e) (f)

Fig. 3. Sketch of scenarios: a) Scenario s_1, b) Scenario s_2, c) Scenario s_3, d) Scenario s_4, e) First part of a possible Scenario s_5 where the test person starts to tie his shoes, f) Second part of a possible Scenario s_5 where the test person directly intents to walk after tieing his shoes.

and linear SVM classification [42]. The frame where a human observer recognizes the initial foot movement, below denoted as *initial frame*, is labeled manually. In order to create a database for training and testing, the pedestrian ROIs are separated into positive and negative examples using the information of the labeled initial frames. In particular, the first 30 frames before the initial frame are the negative examples and the 30 after the initial frame are the positive examples.

Note, however, that application of this hard separation for classifier training could potentially impair classification performance for two major reasons. First, the data closest to the initial frame are hardly discriminative. Second, it is challenging to label the exact frame. To cope with this problem, we discard the 5 adjacent patches of each class to the initial frame, leaving the decision about the correct initial frame to the generalization ability of the SVM. By that means we have splitted the 4663 patches in 2700 positive and 1963 negative examples. For training we use a subset of 1700 positive and 963 negative examples (see Fig. 4). The remaining 1000 examples of each class were used for balanced testing. Training and testing sets used disjunctive video sets.

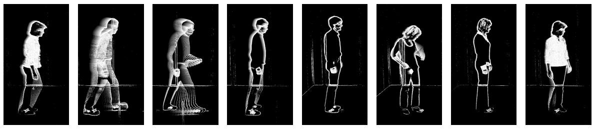


Fig. 4. MHIs of the motion database. Four images on the left: positive examples, on the right: negative examples.

Fig. 5 plots the classification results we achieved. The points of this ROC graph were generated by performing a sweep of the SVM parameter C. The 9 point sets represent different parameter settings, IDs see Table I. We observe maximum true positive rate (99,1 %) at lowest false positive rate (1,5 %) and maximum accuracy (98,8 %) for a cell size of 32 × 32 px and 12 bins.

The elements of the confusion matrix of the best classifier are denoted in the last four columns of Table I. We use the configuration labeled bold (data line 2, minimum false negatives at low false positive rate) as default descriptor and basis in the following evaluation. The associated penalty multiplier of the linear SVM is $C = 2^{-8}$.

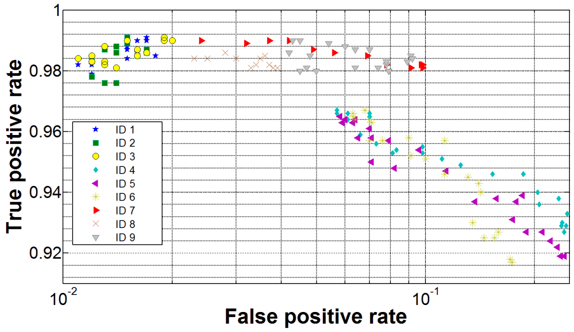


Fig. 5. Receiver operating characteristics (ROC) for MCHOG/SVM classification.

Fig. 6 illustrates the detection of the intention of a pedestrian to start walking by displaying a warning symbol 200 ms after the initial frame. Note, that the motion of the pedestrian, where he obviously discusses with his counterpart before starting (Scenario 5), involving significant motion of the upper extremities, is not detected erroneously.

ID	Cell height (px)	Cell width (px)	Bins	TP rate	TN rate	FP rate	FN rate
1	32	32	9	.991	.983	.017	.009
2	**32**	**32**	**12**	.991	.985	.015	.009
3	32	32	15	.988	.987	.013	.012
4	64	64	9	.967	.943	.057	.033
5	64	64	12	.965	.942	.058	.035
6	64	64	15	.966	.937	.063	.034
7	8	64	9	.990	.976	.024	.010
8	16	32	9	.984	977	.023	.016
9	32	64	9	.990	.957	.043	.010

TABLE I

CLASSIFICATION RESULTS ACCORDING TO PARAMETER VARIATION.

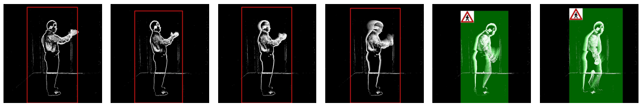


Fig. 6. Example sequence. Note, that the background artifacts do not impair classification.

D. Response Time

The response time of our method is evaluated on a frame-rate-basis. Real-time operation was not in the scope of this study. We do expect that this can be done with hardware acceleration. It was demonstrated that time critical components of the algorithms (MOG, MHI, HOG) can be calculated on the fly using an FPGA for the respective resolution and frame rate [18], [19]. Fig. 7 shows accuracy vs. frame number.

We observe that within 3 - 6 frames on average, i.e. 60 - 120 ms after the manually labeled initial frame, our algorithm detects the movement with an accuracy of 80 %. The accuracy reaches 99 % after 17 frames, i.e. 340 ms, and decreases towards the initial frame. Accuracy is below 100 % before the initial frame because the test persons do not really stand still, whereas after the initial frame spreading of legs is very discriminative for initiation of gait. In relation to typical gait data the result corresponds to a detection within the time span of the first step. It is in the same order of the detection time reported for stopping [6].

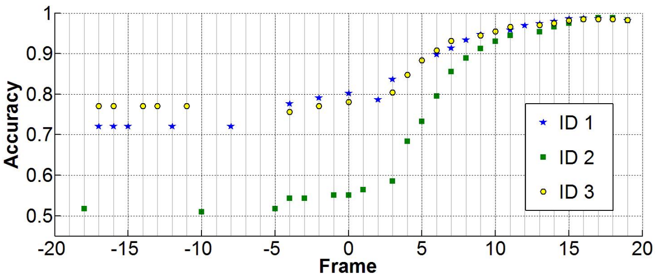


Fig. 7. Accuracy $\frac{TP+TN}{P+N}$ as a function of time. (See also [6], Frame 0 corresponds to the initial frame.)

Fig. 8 presents the positions of head, center of gravity (COG), left and right foot, and the respective velocities during initiation of gait, manually deduced from pixel data of a calibrated camera. Obviously, the COG-data which are widely used for trajectory-based motion estimation are the least sensitive for detection of gait initiation. Video-based recognition additionally induces noise, e.g. due to the scatter of the detected bounding box. Position and velocity of the head, due to body bending (forward in the case of starting and backward in the case of stopping) are slightly and that of the legs are significantly more sensitive. Our MCHOG-descriptor implicitly comprises these features without using noisy absolute positions and its derivatives, respectively.

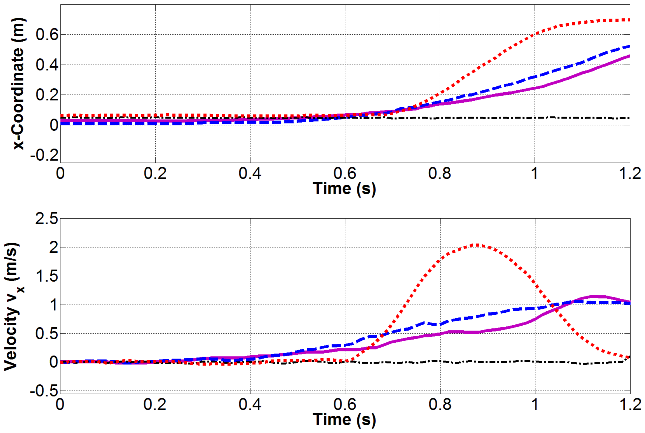


Fig. 8. Position and velocity during initiation of gait. The initial frame is observed at 0.62 s. Left foot (dotted red line) and right foot (dashed black line), head (broken blue line) and COG (solid magenta line).

E. Real World Application

Real world scenarios are more complex, issues such as shadows or illumination changes have to be adressed. To show feasabilty of our method, we process sequences captured at the Ko-FAS [5] test intersections using two high definition cameras mounted 5 meters above street level and looking perpendicular to eachother at the same sidewalk corner, see Fig. 9.

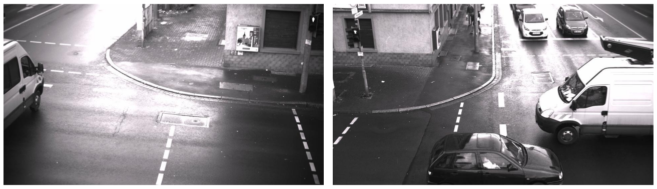


Fig. 9. Images of the Ko-FAS test intersection [5] located at the University of Applied Sciences Aschaffenburg, captured from two different cameras (viewpoint left and viewpoint right).

This camera setup adresses the main conflict scenarios of pedestrians at intersections according to the German In-Depth Accident Study (GIDAS) [43]: a vehicle turns right or left and conflicts with a pedestrian on a crosswalk. The setup ensures the perpendicular view to the pedestrians at the crosswalk independent from the lane to cross. Using MOG background estimation [11] we apply our classifier for gait initiation trained on the laboratory dataset. Fig. 10 illustrates that the proposed method works and responds within a fraction of the first step. The detection of the pedestrian's intention to cross the street is obtained 160 ms after the initial frame in this sequence.

V. Conclusion

In this paper we have proposed a motion contour image based HOG-like descriptor (MCHOG) in combination with an SVM learning algorithm that decide within the initial step if a pedestrian at the curb will enter the traffic lane. The method is designed for monocular-video-based stationary intersection monitoring. By evaluation of ROC and the span of time necessary for recognition, in a case study we demonstrate detection within 120 - 340 ms after the manually labeled gait initiation at accuracy levels of 80 % - 99 %, respectively. The presented data have been evaluated at laboratory conditions and feasibility of the method in a real world intersection scenario is demonstrated. Future work will concentrate on optimization of the response time and on the application to a real test intersection in order to issue a warning to the road traffic. The features described here reflect the person's behavior comprised in the database. To date, there is no reference database for the initiation and termination of human gait in real world urban traffic scenarios.

VI. Acknowledgment

The authors gratefully acknowledge technical support from Continental Safety Engineering, Alzenau, Germany. This work partially results from the joint project Ko-PER, which is part of the project initiative Ko-FAS, and has been funded by the German *Bundesministerium für Wirtschaft und Technologie* (Federal Department of Commerce and Technology) under grant number 19S9022B.

References

[1] "World Health Organization: Global status report on road safety: time for action," 2009.

[2] "U.S. Department of Transportation's (DOT's) IntelliDrive program." Internet: http://www.intellidriveusa.org, Jan. 03, 2011 [Jan. 15, 2011].

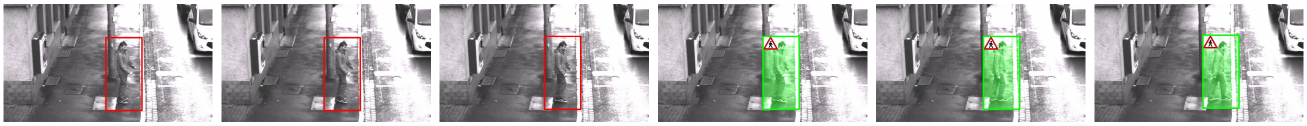


Fig. 10. Example sequence of a detection in the real world application based on the camera with the view of Fig. 9 right.

[3] "SAFESPOT, cooperative vehicles and road infrastructure for road safety." Internet: http://www.safespot-eu.org/, Jul. 23, 2010 [Nov. 04, 2011].

[4] "INTERSAFE-2, cooperative intersection safety." Internet: http://www.intersafe-2.eu, Sep. 02, 2010 [Jan. 15, 2011].

[5] "Ko-FAS, Kooperative Sensorik und Kooperative Perzeption für die präventive Sicherheit im Strassenverkehr." Internet: http://www.kofas.de, [Jan. 15, 2011].

[6] C. Keller, C. Hermes, and D. Gavrila, "Will the pedestrian cross? probabilistic path prediction based on learned motion features," in *Pattern Recognition*. Springer, 2011, vol. 6835, pp. 386–395.

[7] S. Nedevschi, S. Bota, and C. Tomiuc, "Stereo-based pedestrian detection for collision-avoidance applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 380–391, 2009.

[8] D. Pfeiffer and U. Franke, "Efficient representation of traffic scenes by means of dynamic stixels," in *Intelligent Vehicles Symposium (IV), IEEE*, 2010, pp. 217–224.

[9] C. G. Keller, M. Enzweiler, M. Rohrbach, D. F. Llorca, C. Schnorr, and D. M. Gavrila, "The benefits of dense stereo for pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1096–1106, 2011.

[10] S. Wender and K. Dietmayer, "3D vehicle detection using a laser scanner and a video camera," *IET Intelligent Transport Systems*, vol. 2, no. 2, pp. 105–112, 2008.

[11] D. Weimer, S. Köhler, C. Hellert, K. Doll, U. Brunsmann, and R. Krzikalla, "Gpu architecture for stationary multisensor pedestrian detection at smart intersections," in *Intelligent Vehicles Symposium (IV), IEEE*, 2011, pp. 89–94.

[12] T. Gandhi and M. Trivedi, "Pedestrian protection systems: Issues, survey, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 3, pp. 413–430, 2007.

[13] M. Enzweiler and D. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2179–2195, 2009.

[14] D. Geronimo, A. Lopez, A. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1239–1258, 2010.

[15] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5975165&tag=1

[16] S. Maji, A. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Computer Vision and Pattern Recognition, CVPR.*, 2008, pp. 1–8.

[17] C. Wojek, G. Dorkó, A. Schulz, and B. Schiele, "Sliding-windows for rapid object class localization: A parallel technique," in *Proceedings of the 30th DAGM symposium on Pattern Recognition*. Springer, 2008, pp. 71–81.

[18] S. Bauer, U. Brunsmann, and S. Schlotterbeck-Macht, "FPGA implementation of a HOG-based pedestrian recognition system," in *Proceedings of the 42th MPC Workshop*, 2009, pp. 49–58. [Online]. Available: http://www.mpc.belwue.de/Public/WorkshopBaende

[19] J. Kempf, M. Schmitt, S. Bauer, U. Brunsmann, and K. Doll, "Real-time processing of high-resolution image streams using a flexible FPGA platform," in *Embedded World Conference, Nuremberg, Germany*, 2012.

[20] Z. Huijing, C. Jinshi, Z. Hongbin, K. Katabira, S. Xiaowei, and R. Shibasaki, "Sensing an intersection using a network of laser scanners and video cameras," *IEEE Intelligent Transportation Systems Magazine*, vol. 1, no. 2, pp. 31–37, 2009.

[21] S. Bauer, S. Köhler, K. Doll, and U. Brunsmann, "FPGA-GPU architecture for kernel SVM pedestrian detection," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010, pp. 61–68.

[22] Y. Breniere and M. Do, "When and how does steady state gait movement induced from upright posture begin?" *Journal of Biomechanics*, vol. 19, no. 12, pp. 1035 – 1040, 1986.

[23] T. Fugger, B. Randles, A. Stein, W. Whiting, and B. Gallagher, "Analysis of pedestrian gait and perception-reaction at signal-controlled crosswalk intersections," *Transportation Research Record*, vol. 1705, no. 1, pp. 20–25, 2000.

[24] T. Oberg, A. Karsznia, and K. Oberg, "Basic gait parameters: reference data for normal subjects, 10-79 years of age," *Journal of Rehabilitation Research and Development*, vol. 30, no. 2, pp. 210–223, 1993.

[25] R. W. Bohannon, "Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants," *Age and Ageing*, vol. 26, no. 1, pp. 15–19, 1997.

[26] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[27] R. Poppe, "Vision-based human motion analysis: An overview," *Computer Vision and Image Understanding*, vol. 108, pp. 4–18, 2007.

[28] A. Bobick and J. Davis, "An appearance-based representation of action," in *13th International Conference on Pattern Recognition*, vol. 1, 1996, pp. 307–312.

[29] D. Chen and R. Yan, "Activity analysis in privacy-protected video," 2007. [Online]. Available: http://www.informedia.cs.cmu.edu/documents/T-MM_Privacy_J2c.pdf

[30] L. Jianyi and Z. Nanning, "Gait history image: A novel temporal template for gait recognition," in *IEEE International Conference on Multimedia and Expo*, 2007, pp. 663–666.

[31] C. C. Lee, C. H. Chuang, J. W. Hsieh, M. X. Wu, and K. C. Fan, "Frame difference history image for gait recognition," in *International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 4, 2011, pp. 1785–1788.

[32] M. A. R. Ahad, J. K. Tan, H. Kim, and S. Ishikawa, "Motion history image: its variants and applications," *Machine Vision and Applications*, pp. 1–27, October 2010.

[33] C. E. Rotgers, D. F. Greenlee, and R. D. Blomberg, "System and method for providing pedestrian alerts," Patent US 7 095 336, 2006.

[34] M. Farmer, L. H. Rein, and A. Jain, "Interacting multiple model (IMM) Kalman filters for robust high speed human motion tracking," in *Proceedings 16th International Conference onPattern Recognition*, vol. 2, 2002, pp. 20–23.

[35] S. Schmidt, B. Färber, and A. Pèrez Grassi, "Geht er oder geht er nicht? - ein FAS zur Vorhersage von Fussgängerabsichten," in *5. Workshop Fahrerassistenzsysteme*. Freundeskreis Mess- und Regelungstechnik Karlsruhe e.V., 2-4 April 2008, pp. 176–184.

[36] S. Schmidt and B. Färber, "Pedestrians at the kerb – Recognising the action intentions of humans," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 4, pp. 300–310, 2009.

[37] S. Bauer and S. Zecha, "Erkennung von Personen, insbesondere von Fussgängern," Patent DE102 008 062 915A1, 2008.

[38] C. G. Keller, T. Dang, H. Fritz, A. Joos, C. Rabe, and D. Gavrila, "Active pedestrian safety by automatic braking and evasive steering," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1292–1304, 2011.

[39] N. Shiozawa, S. Arima, and M. Makikawa, "Virtual walkway system and prediction of gait mode transition for the control of the gait simulator." *International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, pp. 2699–2702, 2004.

[40] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999, pp. 246–252.

[41] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *17th International Conference on Pattern Recognition, (ICPR'04)*, vol. 2, 2004, pp. 28–31.

[42] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, CVPR*, vol. 1, 2005, pp. 886–893.

[43] "German in-depth accident study." [Online]. Available: http://www.gidas.org