2010 IEEE Intelligent Vehicles Symposium
University of California, San Diego, CA, USA
June 21-24, 2010

Pedestrian Crossing Detecting as a part of an Urban Pedestrian Safety System

Sebastian Sichelschmidt, Anselm Haselhoff, Anton Kummert
Faculty of Electrical, Information and Media Engineering
University of Wuppertal, Germany
{sichelschmidt,haselhoff,kummert}@uni-wuppertal.de
Martin Roehder, Björn Elias
Audi Electronics Venture GmbH
Gaimersheim, Germany
{martin.roehder,bjoern.elias}@audi.de
Karsten Berns
Robotics Research Lab
University of Kaiserslautern, Germany
berns@informatik.uni-kl.de

Abstract— **Although recent statistics demonstrate a decrease in pedestrian fatalities, the absolute number of accident related deaths is sufficiently high to justify research in the area of vulnerable road user protection. Research has shown that situation awareness, which requires a significant amount of context information, is critical to timely intervention. Altered pedestrian behavior, due to traffic regulation, requires context information. For example, crosswalk presence and location knowledge can be of importance in pedestrian crossing scenarios. Therefore this paper discusses the implementation of a crosswalk detection algorithm, using Fourier transformation, augmented bipolarity, inverse perspective mapping template matching and edge orientation ratios for classification.**

I. Introduction

Regarding today's traffic injury statistics, many pedestrians still die especially in urban scenarios. Due to this, guidelines and legislation focus more and more on automotive safety systems developed to protect pedestrians. Meanwhile a lot of car manufacturers announce forward looking safety systems designed primarily to mitigate injuries of vulnerable road users, but the long term goal is to prevent accidents.

In order to avoid a collision, an action such as braking or steering has to be initiated. The dilemma is the safety system has to intervene early and sometimes even at a time, a pedestrian could still mitigate the situation. Therefore, a complex situation analysis is needed to create a reasonable basis for the specific decision and to decrease false alarms to a minimum. In reference to GIDAS (German In-Depth Accident Study), a very detailed traffic injury statistic, the majority of pedestrian related accidents occurs in pedestrian crossing scenarios [1]. In these scenarios a car is driving straight and a pedestrian crosses the street. It can appear in the context of no traffic regulation, a traffic light handling crossing permissions and a pedestrian crosswalk enhancing pedestrian privileges. Every context type creates a different situation which has to be handled differently by a safety system.

In common driver assistance or active safety systems, a sensor perceives a car's environment, a processing unit processes the sensor data and classifies the situation, and a decision level unit determines an appropriate action. The situation classification itself is very challenging because it requires environmental data which probably does not comply with the unprocessed sensor data or, more likely, which demands more than one sensor. As a consequence, one has to interpret the environmental sensor data first before it becomes possible to gather higher level information which allow situation indication. To achieve robust situation classification, a lot of context information is needed [2]. This paper introduces the detection of crosswalks as one important part of an urban pedestrian safety system.

Section II summarizes current activities in the area of crosswalk detection. Section III describes how to extract the crosswalk context from an imaging sensor. In section IV we present current detection rates for urban scenarios. Section V concludes this paper and provides an outlook on future work.

II. State of the Art

The detection of pedestrian crossings - commonly known as zebra crosswalks - has been implemented in very few applications so far. The most complex and detailed approach has been developed in projects for vision aids like MOVIS[1] for instance. In early approaches, a grouping algorithm, based on projective geometry constraints [3], was augmented by a model based search [4]. One interesting feature related to this topic is called bipolarity [5]. It determines the

[1]Mobile Optoelectronic Visual-Interpretative System for the Blind and Visually Impaired

978-1-4244-7868-2/10/$26.00 ©2010 IEEE

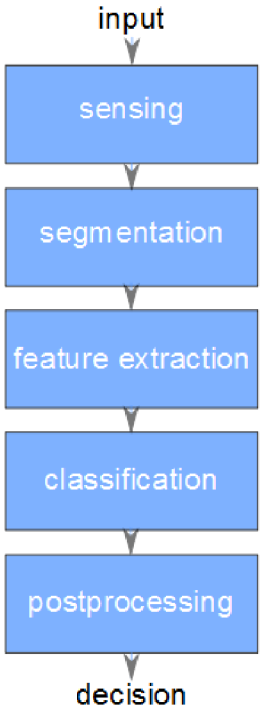


Fig. 1. Pattern recognition chain [6]

distribution of black and white pixels inside a certain area of an image. In a zebra crossing section of an image black and white pixels have almost equal distributions.

Implementing a pedestrian crossing detection in a vehicle has not been discussed so far. The advantage in comparison to the above-named vision aids is a lower degree of freedom. Due to the knowledge of the positioning of the camera with respect to the car and the street, assumptions about the crossing's orientation can be made. The downside however is the higher speed of a car. This leads to the necessity of developing a detection with much longer range.

Because of a zebra crossings planar nature no 3D-sensors like a Photonic Mixer Device (PMD) - which have been established in the automotive environment - can be used for its detection[2]. Therefore the camera remains the most useful sensor. Respecting the OEM's and supplier's strict price policy the most basic attempt should be done on a gray scale mono video camera system as presented in this paper.

III. Detection of pedestrian crossings

The detection of pedestrian crossings will be regarded as a standard problem of pattern recognition as specified in [6]. The classical workflow is displayed in figure 1.

A. Sensors

A gray scale camera has been placed and calibrated inside the car near the rear view mirror.

[2]Allthough those sensors might play a role in measuring the pedestrian crossings distance

B. Segmentation

Before features can be extracted a hypothesis has to be generated by applying image processing algorithms. These steps are detailed in the following enumeration

1) Pitch angle compensation: Driving on a non ideal surface leads to roll, yaw and most notably pitch movements of the car and thus to the need of resetting the camera's extrinsic parameters. Otherwise the estimation of distances will be highly inaccurate on greater distances (and a long range is a fundamental specification of this system, as stated above). Information about these angle alterations can be either retrieved from available or additional vehicular sensors or extracted directly from the picture via movement-detection-algorithms. The results of this paper have been accomplished without any angle compensation, it is recommended to use an accurate pitch compensation nevertheless.

2) Region of interest: To approximatively get an image just of the street, the camera image must be bordered. Equation 1 extracts a pixel function $g(u,v)$ from the original image $f(u,v)$ (with $0 \leq u < M$ horizontal resolution and $0 \leq v < N$ vertical resolution) without parts that are either not on ground level (above horizontal line j_{ground}) or which cannot belong to the road surface because of perspective distortion.

$$g(u,v) = \begin{cases} 0 & \text{if } v < j_{\text{ground}} \\ 0 & \text{if } u + 2v - j_{\text{ground}} < \frac{M}{3} \\ 0 & \text{if } u - 2v - j_{\text{ground}} > \frac{2M}{3} \\ f(u,v) & \text{else} \end{cases} \quad (1)$$

A detection of the road boundaries would lead to improved results if it is able to detect the road boundaries even in urban environments.

3) Histogram equalization: To support subsequent edge detection, the image's contrast is enhanced by applying a histogram equalization.

4) Edge extraction: The rectangular shape of the crossing's stripes allows edge detection algorithms to be implemented. A Canny approach is suited best, because of its edge-width reduction to one pixel and its adaptable threshold hysteresis.

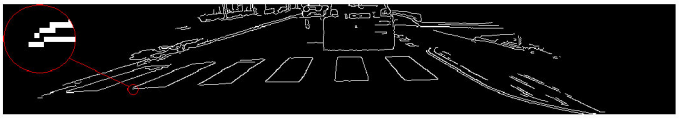
5) Dilation: To avoid disjointed edges rather than a closed contour the morphologic operation of dilation is applied to the binary edge image.

6) Contour extraction: In an ideal world with perfect zebra crossing stripes and non shaking cars it would be easy to look for rectangular forms using Inverse Perspective Mapping (IPM) [7] [8]. In real life however results are better by simply searching for closed contours and "filling" them in a first step before applying a FFT (see below).

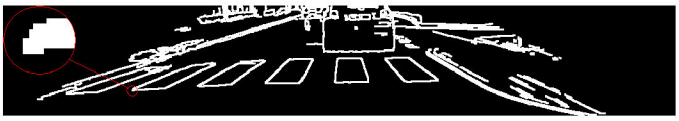
7) Fourier-Transform: Given the real world distance of each image row using the camera calibration and information about the zebra stripes width - which is assumed to be the same as the clearance between the stripes - there is a particular binary pattern which can be searched for. Since this pattern can be expressed as a frequency, a Discrete Fourier Transform is applied to each image row. Using only the contour lines would lead to a Dirac impulse sequence

(a) input image

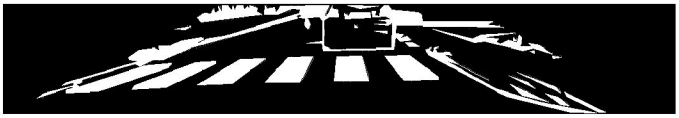
(b) region of interest



(c) canny filtered binary image

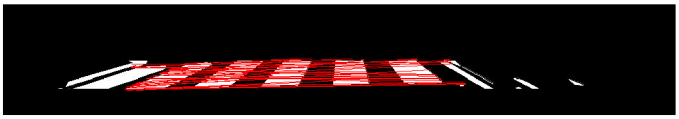


(d) dilated image

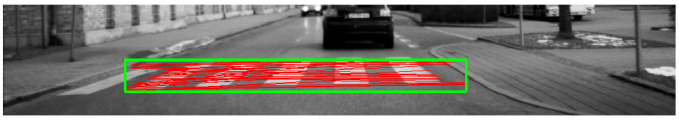


(e) filled contours

(f) FFT filtered image



(g) hough lines



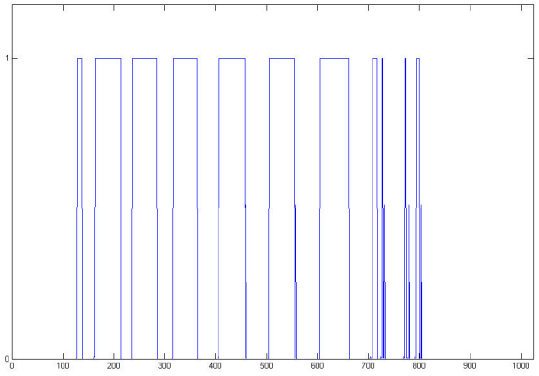
(h) bounding box

Fig. 2. States of segmentation

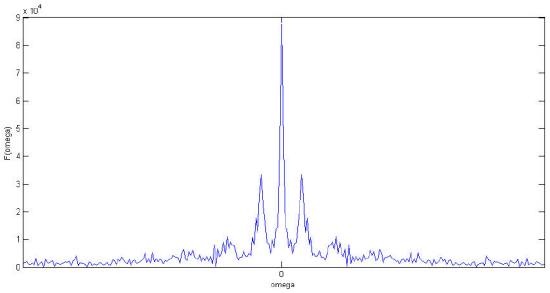
with the frequency f_{zebra} regarding to the stripe contours as a function of the distance. This however would be most sensitive to noise, that is why the contours have been filled in the previous step. This leads in the ideal case to a rectangular function $f(u) = \text{rect}(u)$ and the Fourier transformed spectrum

$$F(\omega) = \mathcal{F}\{\text{rect}(u)\} = |T|\,\text{sinc}(\frac{\omega T}{2}). \tag{2}$$

Figure 3 shows line 110 (of 167) of figure 2(e) and its spectrum. The secondary lobes can be seen clearly and are



(a) row signal



(b) linear row spectrum

Fig. 3. FFT

shifted by

$$\pm\omega_{\text{zebra}} = \pm M \cdot f_{\text{zebra}} = \pm\frac{M}{T_{\text{row}}} \tag{3}$$

with

$$T_{\text{row}} = \frac{\frac{M\,[\text{pix}]}{2} T_{\text{real}}\,[\text{m}]}{d\,[\text{m}]\,\tan(\frac{\alpha}{2})} \tag{4}$$

where T_{real} is the period length and d the row's distance in the real world and α the camera's angle of aperture. Due to noise a small tolerance should be permitted.

8) Hough Lines: To exclude edges that do not belong to the pedestrian crossing from the FFT-filtered image, a Hough transform [9] is applied and lines with angles in the range of $\theta \approx 0°$ are extracted. Finally only these lines are shown in figure 2(g), which are continued with a certain length after a gap of roughly the same length. A bounding box spanning these lines is highly suited as a hypothesis for the following feature extraction.

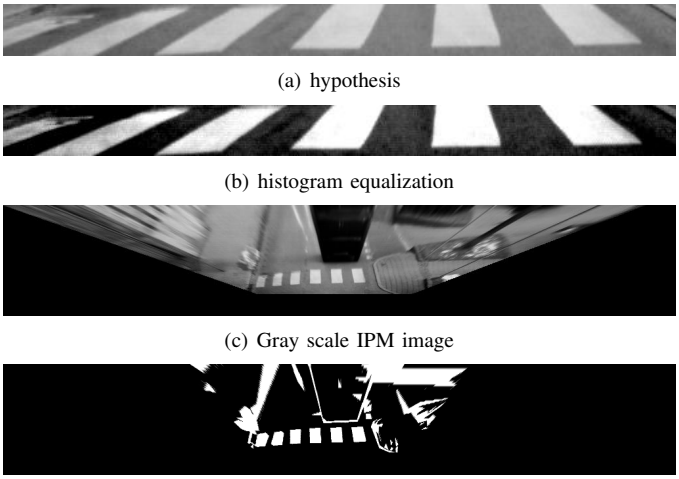
C. Feature extraction

The next step is the implementation of certain predefined highly variant features, each resulting in a so called feature vector spanning the feature space. The following features have been embedded.

1) Fourier Transform: This feature works similar to the FFT in the segmentation section. Here however, instead of a simple comparison to a certain threshold, the median of the highest secondary lobe within the defined tolerance around

the expected frequency in either the ROI (figure 2(b)) and the contour image (figure 2(e)) is returned.

2) Augmented bipolarity: The already mentioned bipolarity [5] has been defined for the binary case that is an equal amount of black and white pixels. Since not all contours can be detected at any given time, due to occlusion or bad marking condition, this feature does however not lead to satisfying results. Therefore the boundary between 'zero' and 'one' has been extended to a histogram examination. The gray scale ROI image (figure 4(a)) is first equalized (figure 4(b)) followed by the element-wise allocation to four histogram bins. The thresholds may be set either statically or dynamically. For the latter one may choose between the mean or median of all intensity values and subsets or an application of Otsu's method [10]. The features emerge from the proportion of the lower two bins compared to the total number of pixel and the middle two bins likewise. Each feature for itself does not contain much information. In combination however a value between 0.5 and 0.7 for the first feature and a low value (< 0.4) for the second feature indicate a pedestrian crossing.

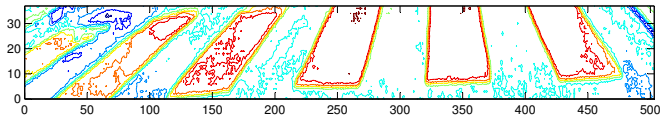


(a) hypothesis

(b) histogram equalization

(c) Gray scale IPM image

(d) Binary IPM image



(e) gradient image

Fig. 4. Feature extraction elements

3) IPM template matching: The formerly discussed Inverse Perspective Mapping proved itself to be well applicable for a template matching. On both, a gray scale (figure 4(c)) and a binary contour IPM image (figure 4(d)) two slightly differently rotated templates with three stripes in five different scales are shifted over the hypothesis region. The highest coefficients of correlation in each image represent the two features of this method.

4) Edge orientation ratio: The orientations of edges in an image section containing a pedestrian crossing hold distinctive cumulations. Defining gradients in u- and v-direction as

$$\nabla f_u(u,v) = \frac{\delta f}{\delta u} \tag{5}$$

$$\nabla f_v(u,v) = \frac{\delta f}{\delta v}, \tag{6}$$

the angle ϕ between u-axis and the edge can be written as

$$\phi = \arctan\left(\frac{\nabla f_v(u,v)}{\nabla f_u(u,v)}\right) + 90^\circ \tag{7}$$

with an amplitude of

$$|\nabla f| = \sqrt{(\nabla f_u(u,v))^2 + (\nabla f_v(u,v))^2}. \tag{8}$$

The amplitudes can be seen in figure 4(e), while the number of edges with angle ϕ and weighted with the gradient's amplitudes are displayed in figure 5. The cluster in the area

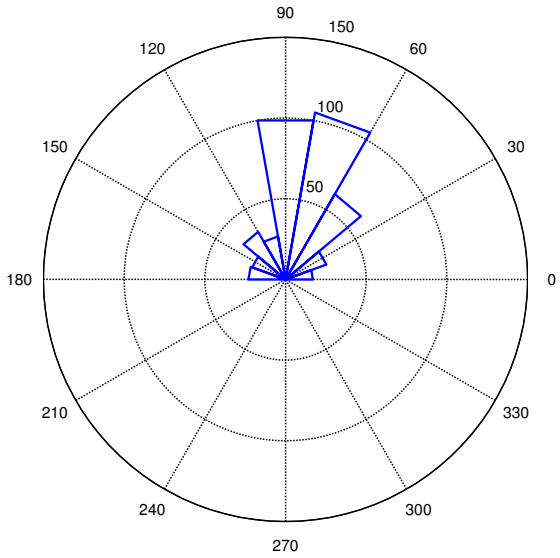


Fig. 5. Edge orientation (ϕ) histogram weighted with the gradient's amplitudes

$60^\circ < \phi < 100^\circ$ is typical for right-hand traffic and therefore poses a good feature. Another contribution to the feature space states the small value of the combined $0^\circ < \phi < 20^\circ$ and $160^\circ < \phi < 180^\circ$ bins.

D. Classification

A classifier sets a decision boundary into the feature space, so that future hypotheses and their feature vectors can be assigned to one of the defined classes. In this case a (Gaussian Kernel) Support Vector Machine (SVM) has been used with supervised learning and stratified cross validation. The results can be seen in section IV.

E. Post processing

As in most cases the classification is not free of error. Especially so called false positive declarations - detected crossings that do not exist - may flash up from time to time.

Rate	Value
TPR	92.34%
TNR	75.13%
PPV	69.82%
NPV	94.50%
Y_{weighted}	0.88

TABLE I
CLASSIFICATION RESULTS

These could be eliminated with a tracking algorithm like a Kalman filter for instance. This however leads to a generally delayed detection which might compensate the advantage of fewer errors.

IV. CLASSIFICATION RESULTS

A classifier's performance can be measured in different ways. Quite established are parameters like the sensitivity -or true positive rate (TPR)- and the specificity -or true negative rate (TNR)- which describe the probability that an object is assigned to the correct class. In addition the positive predictive value (PPV) and the negative predictive value (NPV) describe the probability that any object in a certain class has been classified correctly. Another meaningful parameter poses the so called weighted Youden-Index (Y_{weighted}) [11] which is a combination of TPR and TNR, stressing one of these two values.

$$Y_{\text{weighted}} = m\frac{t_p}{t_p + f_n} + (1-m)\frac{t_n}{f_p + t_n} \qquad (9)$$

Since the classifiers task is to save the life of pedestrians it is reasonable to put weight on the TPR and the NPV. With $m = 0.75$ this leads to the results shown in table I. In this case a training and validation feature vector set of 1594 positive and 2610 negative examples were used. This training set contains all kind of different road and weather conditions, such as paved roads and snow-covered ground for instance. It is therefore universally applicable though the overall detection rate is naturally lower than that of an algorithm which does not cover all these special cases. The algorithm shows quite a good robustness against partial occlusion though.

Further analysis shows that nearly 80% of the system's runtime is generated during the template matching. Therefore it might be advisable to exclude this feature from the feature space which leads to relatively small expenses ($Y_{\text{weighted}} - 0.05$) in the classifiers quality.

The system range depends on different exterior parameters. The authors were able to achieve a system range between 14.1 and 32.0 meters (average 22.9 meters) which is enough to perform a change maneuver or even to stop the car in an urban environment.

V. CONCLUSION

In this paper the importance of context information for an active pedestrian safety system has been highlighted. Pedestrian crossing scenarios are focused on, because they were dominant in a detailed traffic accident statistic. According to this scenario, pedestrian crossing detection based on gray scale images is introduced. The different processing steps like segmentation, feature extraction and classification have been described in detail. The results of an extensive evaluation including different weather conditions have been presented.

Interpreting the results, a pedestrian crossing detection based on gray scale images is generally possible. The usability of this information in a situation analysis strongly depends on a confidence parameter which is necessary to know the reliability of the information at each time. It can be derived from an observation over time or by evaluating single feature votes for a certain class.

Further experiments will show the impact on the precision of predicting pedestrian behavior in crossing scenarios.

REFERENCES

[1] M.-M. Meinecke, M. A. Obojski, E. Wykowski, M. Stanzel, H. Riedel, B. Elias, W. Branz, D. Gavrila, A. H, G. Baratoff, S. Munder, J. Vetter, "Unfallanalyse und Funktionsspezifikation - AKTIV-SFR", 2007

[2] M. Roehder, B. Elias, B. Giesler, K. Berns, "Predicting pedestrian behavior in an automotive safety system", Workshop on Intelligent Transportation, Hamburg, Germany 2009

[3] S. Utcke, "Grouping based on Projective Geometry Constraints and Uncertainty", Internal Report 1/97 Technische Universitaet Hamburg Harburg Technische Informatik I, 1997

[4] S. Se, "Zebra-crossing Detection for the Partially Sighted", Proceedings IEEE International Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, USA, 2000

[5] M.S. Uddin, T. Shioyama, "Bipolarity- and Projective Invariant-Based Zebra-Crossing Detection for the Visually Impaired", 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005

[6] R.O. Duda, P.E. Hart, D.G. Stork, "Pattern Classification, 2nd edition", John Wiley & Sons, Inc., 2001

[7] H.A. Mallot, H.H. Buelthoff, J.J. Little, S. Bohrer, "Inverse perspective mapping simplifies optical flow computation and obstacle detection", Biological Cybernetics, 64, 177-185, 1991

[8] A.M. Muad, A. Hussain, S.A. Samad, M.M. Mustaffa, B.Y. Majlis, "Implementation of Inverse Perspective Mapping algorithm for the development of an automatic Lane Tracking System", IEEE 0-7803-8560-8/04, 2004

[9] P.V.C. Hough, "Methods and Means for Recognizing Complex Patterns", U.S. Patent 3,069,654, 1962

[10] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, 1979, pp. 62-66, 1979

[11] W.J. Youden, "Index for Rating Diagnostic Tests", Cancer. 3. 1950, 3235, 1950