Pedestrian Detectability: Predicting Human Perception Performance with Machine Vision

David Engel and Cristóbal Curio
Max Planck Institute for Biological Cybernetics
72076 Tübingen, Germany
david.engel@tuebingen.mpg.de
cristobal.curio@tuebingen.mpg.de

***Abstract*— How likely is it that a driver notices a person standing on the side of the road? In this paper we introduce the concept of pedestrian detectability. It is a measure of how probable it is that a human observer perceives pedestrians in an image. We acquire a dataset of pedestrians with their associated detectabilities in a rapid detection experiment using images of street scenes. On this dataset we learn a regression function that allows us to predict human detectabilities from an optimized set of image and contextual features. We exploit this function to infer the optimal focus of attention for pedestrian detection. With this combination of human perception and machine vision we propose a method we deem useful for the optimization of Human-Machine-Interfaces in driver assistance systems.**

I. INTRODUCTION

In this paper we present the concept of pedestrian detectability which measures the probability that a pedestrian is detected by an observer 'at a glance'. This concept can allow driver assistance systems to estimate which pedestrians are unlikely to have been noticed by the driver, hence posing a greater risk for collision. We present a novel machine-vision approach for predicting the detectability of pedestrians in natural scenes and demonstrate that knowledge about the characteristics of the detectability of pedestrians can improve user-performance in a localization task.

The main goal of driver assistance systems is to make traffic safer for all participants (*e.g.* by providing feedback about possibly dangerous situations or even by active intervention for collision avoidance). Collisions between cars and pedestrians are a major source of danger. According to the BAST [32], there were 31.647 traffic accidents that involved pedestrians in Germany in 2009. Consequently, driver assistance systems usually aim at detecting pedestrians and informing the driver about possible risks. Over the course of the last several years, pedestrian detection and tracking has constituted an active field of research and major advances have been achieved (*e.g.* [12], [16], [25]). In spite of their impressive results, current approaches are still not able to match human performance. At first glance, this seems to undermine the usefulness of driver assistance systems, since the computer would have to rely on worse input data than the driver, but would have to accomplish a better prediction of possible risks. Nonetheless, the utility of artificial vision systems can be justified, in the case of distracted drivers or in conditions with bad visibility. Taking into account other modalities, such as infrared cameras and depth sensors, it is reasonable to expect computer vision-based pedestrian detection to outperform humans in the near future.

This work was partly sponsored by EU FET-Open project TANGO (ICT-2009-C 249858).

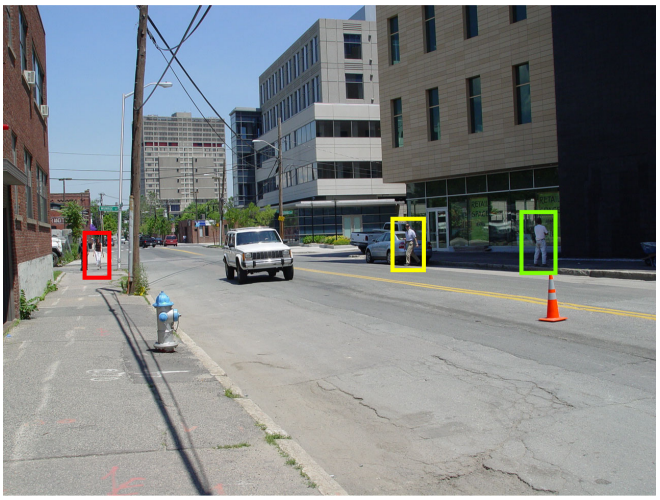

Fig. 1. The notion of detectability measures the ease with which pedestrians can be detected in an image. The pedestrian on the right (green) is easily detected, the one in the center (yellow) is harder to detect because of the difficult lighting conditions. The marked pedestrian (red) on the left has a very low detectability. Predicting these detectabilities allows a driver assistance system to better estimate which pedestrians might be in danger.

Finding the pedestrians that are relevant for the driving task is crucial. However, it constitutes only one part of a useful driver assistance system. An open question from the field of Human-Machine-Interfaces (HMI) is how information is best presented to a driver. The data should be presented in a minimally intrusive fashion while still offering all information that is relevant to the task, thus allowing the driver to avoid collisions without distraction. For example, drawing attention to all the pedestrians in the field of view of the driver by, for example, highlighting their outlines with a head-up display would probably be more distracting than useful. Thus, the following question arises: Which location should the attention of the driver be directed to by a driver assistance system? Ideally, the system should perform a risk assessment and estimate the pedestrians in the given scene that are most at risk. Risk assessment is a difficult problem to solve algorithmically, but it is one that can be done quite well

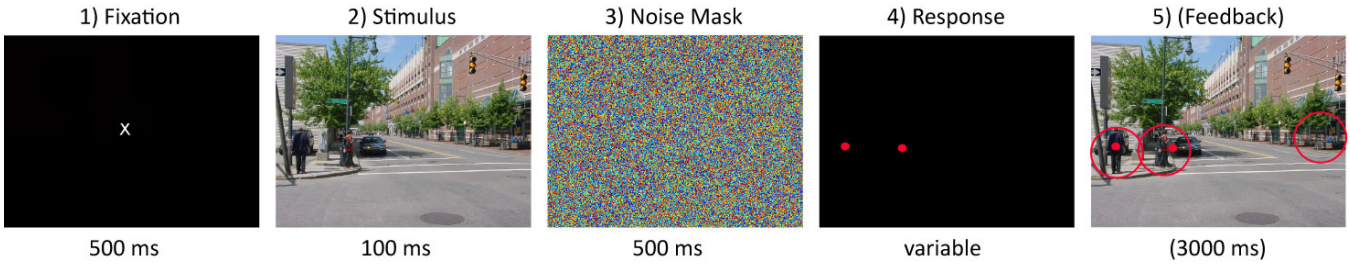

Fig. 2. The design of the experiment to determine the detectabilities of pedestrians. A fixation cross was presented for 500ms (1), followed by the test image from the StreetScenes database for 100ms (2) and a noise mask for 500ms (3) to suppress all further low-level processing. During the response phase (4), participants clicked on all pedestrians they had noticed in the image. In one fifth of the trials, feedback (5) about the true positions of all labeled pedestrians in the image was provided for 3 seconds.

by human drivers. We propose that pedestrians who have not been noticed by the driver present higher risks of collision and that the driver should be alerted to their presence by the assistance system. Consequently, we developed a method to predict which pedestrians have most likely been missed by the driver using methods from computer vision and machine learning. Our framework is able to predict the optimal point where the attention of the driver should be drawn in a sudden hazard situation (e.g. using a head-up display) to maximize the chance that all pedestrians in the scene will be detected.

First, Section II will present some related work on the problem. Section III outlines our method used to create a database of pedestrians along with their associated detectabilities. Section IV shows how this database is used to learn a mapping to predict detectabilities. Finally, Section V demonstrates how the output of such a regression function can be used to successfully predict the optimal fixation point for maximizing human pedestrian localization performance. We discuss our approach in Section VI and provide an outlook on our work.

II. RELATED WORK

To the best of our knowledge, this study is the first one proposing an approach that estimates the detectabilities of pedestrians for HMI-optimization, using a computer vision approach. Most closely related is the recent work by Pomarjanschi *et al*. [31] who used gaze tracking during a driving task in a virtual reality setting to estimate the pedestrians that were likely to have been missed by the driver. Their study, however, neglected the visual properties of the pedestrian appearance and its scene context, which are likely to influence the detectability of pedestrians. Doshi and Trivedi [13] introduced an approach combining gaze tracking with analysis of the environment, using computer vision to predict driver attention. Similarly, Fletcher *et al*. [17] have investigated driver gaze tracking in the context of driver assistance systems. Based on a related idea, Spain and Perona [35] measure and predict the 'importance' of objects in an image, *i.e.* , the order in which they will be named by a human observer. Pinneli *et al*. [30] use a Bayesian framework to predict the perceived 'interest' of an object using various factors such as location, contrast and color.

Most approaches to driver assistance systems address the problem of collision avoidance by detecting (*e.g*. [12], [26], [27], [29]) and tracking pedestrians (*e.g*. [2], [34]). Furthermore, pedestrian pose estimation, which can yield further valuable information about the future actions and paths of persons, has received considerable attention from the computer vision community lately ([1], [28]). Moreover, perceiving and interpreting behavior of groups of pedestrians correctly is a promising research direction, especially in driver assistance scenarios [23]. Integrating such high-level information with the results from our approach could drive an internal simulation of the driver assistance system that predicts the future risks of the current situation and the pedestrians therein (*c.f*. research regarding risk horizon estimation by *e.g*. Laugier *et al*. [18], [24]).

III. MEASURING DETECTABILITY

In order to train an algorithm that is able to estimate human detectabilities from labeled images, we first needed a dataset containing pedestrians with their associated detectabilities. Humans are almost perfect at finding people in images when there are no time constraints, but this does not imply that all pedestrians are equally easy to find in an image. In a driver assistance context, it cannot be assumed that the driver will always devote his full attention to searching for pedestrians. Consequently, the definition of the detectability of a pedestrian we suggest here, is the probability that the position of a pedestrian in an image can be reported correctly by a human observer after the image has been presented for only 100 milliseconds (see Equation 1):

$$\mathcal{D}(\text{Pedestrian}) = p\left(\begin{array}{c}\text{Pedestrian position is}\\ \text{reported after 100ms}\end{array}\right) \quad (1)$$

We opted for the relatively brief presentation time to ensure that only very little high level cognitive processing is taking place and to prevent eye movements (saccades) from having an impact on the perception of the image. Usual saccade latencies are about 200ms (see [10]) and even ultra-rapid saccades (*e.g*. [19], [21]) have a latency of 80ms-100ms and a duration of over 50ms which is considerably longer than the stimulus presentation time chosen here. This definition of detectability captures the idea of the ease of detecting a pedestrian in an image 'at a glance'. Our working

hypothesis is that detectability is highly correlated with the probability that a distracted driver (or one that is not paying full attention to the street) will overlook the pedestrian, which are situations wherein a driver assistance system should step in and alert the driver appropriately.

A. Data

We performed a psychophysical experiment to obtain a set of pedestrians with their associated detectabilities $\mathcal{D}(\text{Pedestrian})$'s. To ensure validity in a relevant setting, we chose a dataset that contains labeled pedestrians in a natural setting. The MIT StreetScenes dataset [6] is well suited for the task since it contains labeled pedestrians in a wide variety of poses and contexts as well as dense labels for other object classes such as cars or sidewalks which might influence the detectability of pedestrians. Furthermore, it includes a large number of images containing two or more pedestrians, which is important since we assume that a higher number of distracting pedestrians might reduce the overall detectability. The StreetScenes database contains several images with suboptimal or ambiguous labelings. However, these are not detrimental in our case since we only needed a subset of a few hundred images from the whole dataset (more than 3500 images) for our experiment. This enabled us to select images with high quality annotations for the experiment by hand.

B. Experimental Design

For the experiment we selected a total of 626 images from the StreetScenes database. 142 contained no pedestrians, 245 contained exactly one pedestrian and 239 contained two or more pedestrians. One trial of the experiment consisted of the following stages: 1) The participants were shown a fixation cross for 500ms. 2) The stimulus image from the database was presented for 100ms. 3) A random noise mask was shown for 500ms to prevent any further low-level processing taking place after stimulus presentation. 4) A black screen was then presented and the participants had to indicate where in the image they perceived pedestrians by clicking on their locations on the screen with a mouse. A red dot appeared where they clicked. 5) In 20% of the trials, chosen at random, the participants were shown the same test image again. This time, it was a composite of the original image and the ground truth positions of the pedestrians and the responses of the participants. This type of feedback was given to allow the participants to correct for any biases in their responses (without feedback, users demonstrated a tendency to click closer to the center of the image). A schematic representation of the experiment for obtaining human detectability characteristics is shown in Figure 2. The experiment was programmed in MATLAB™ using the freely available Psychophysics Toolbox 3 ([8], [22]) to ensure accurate timing.

The rapid presentation intervals can make this an exhausting task for the participants. To avoid fatigue effects, we forced the users to take a short break every 100 trials. The head position of the participants was fixated 65 centimeters in front of the monitor with a chinrest, resulting in a horizontal viewing angle of approximately 60°. An extensive introduction and training phase preceeded the experiment to familiarize the participants with the setup and to minimize response biases. After the experiment, all participants answered a questionnaire. A total of 11 subjects (mean age: 26.4; 7 males, 4 females) participated in the experiment. Of these 11 participants, 10 were right-handed and one was left-handed. A response took 1.5 seconds on average.

Fig. 3. Examples of correct (green crosses) and incorrect (red dots) participant responses in the detectability experiment. The red circle shows the 100 pixel radius around the center of the pedestrian that represents the 'hit zone'. All clicks in that circle were treated as correct detections of that pedestrian. For large pedestrians (left), the whole body also counted as 'hit zone'. In case of multiple pedestrians that are close together (right), ambiguous clicks can occur. Since we are not able to determine post-hoc, which pedestrian was really detected by the user we treat both pedestrians as being 'detected'.

Our response method asks the participants to click on the positions where they have seen a pedestrian from memory. This pointing task is prone to several kinds of noise such as: manual imprecision during clicking, inaccurate memory encoding and retrieval as well as other effects. To compensate for this, we counted every click within a radius of 100 pixels around the center of gravity of a pedestrian as a correct detection. For large pedestrians, we also accepted clicks on the annotation polygon as a hit (see Figure 3). By averaging over all participants' responses, we obtained the detectability characteristics denoted by $\mathcal{D}(\text{Pedestrian})$ for each pedestrian.

C. Results

The average detectability across all pedestrians in our database is 62.97%. A more detailed distribution of detectabilities is shown in the histogram in Figure 4. Broadly speaking, all possible detection probabilities were evenly represented in our database (except for a high number (172) of samples that were correctly detected by all participants). An analysis of the correlations between the percentage of correctly marked pedestrians with the answers given in the questionnaire is summarized in Table I. Only the correlation with driving experience is significant at the 5% confidence level. The correlation with the total number of clicks in all images approaches significance ($p = 0.056$). It is interesting to note that neither gender, experience with video games or the estimated percentage of correct answers, as reported by the participants, showed any significant correlation with their performance (correlations of 0.15, 0.16 and 0.10, respectively).

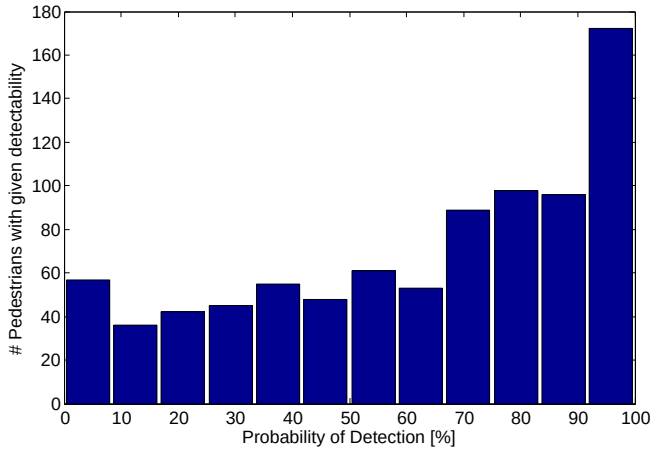

Fig. 4. Histogram of the count of pedestrians for the different detection probabilities. There is a high number of pedestrians (172) that were correctly reported by all participants.

Age	-0.28
Response Time	0.39
Proficiency with Computer	-0.24
Regularity of driving	0.64
Estimated % correct	0.10
Estimated % of images with more than one pedestrian	0.42
Total # of clicks	0.59
Proficiency with video games	0.16
Gender	0.15

TABLE I

CORRELATIONS BETWEEN THE RESPONSES FROM QUESTIONNAIRES AND PEDESTRIAN DETECTION PERFORMANCE.

Name	Description
pHoG	Pyramidial Histogram of Oriented Gradients Descriptor as described in [7]
Pos	Position, Size, Color and Standard deviation of the pedestrian
Area	
PedMean	
PedStd	
PedCount	Total number of pedestrians in the image
DiffMean	Difference in mean color, standard deviation and the earth mover distances between difference types of histograms between the bounding box of the pedestrian and its context (see Figure 5))
DiffStd	
DiffHist	
DiffRGBHist	
DiffLABHist	
Dist2Center	Distance from the center of the pedestrian to the center of the image, to the center of the closest other pedestrian and car in the image
Dist2Ped	
Dist2Car	
PixelPerClass	Number of pixels in the image of each of the eight annotated classes the three forground classes
PixelPerFG	
FixColor	Mean brightness of a 15×15 area around the fixation point in the image and the difference between the fixation point and the mean color of the pedestrian
DiffFix	
mfThres	After resizing the image in the bounding box around the pedestrian to 100×50, we computed the flux flow $\mathcal{F}$ as described in [15]. *mfThres* is the number of pixels whose flux flow is above a threshold of 1 and represents the level of symmetry in this area. *mfCount* is the number of Interest points on the pedestrian. *mfMaxScale* and *mfMeanScale* are the largest and average local scales at the interest points.
mfCount	
mfMaxScale	
mfMeanScale	

TABLE II

NAME AND A SHORT DESCRIPTION OF THE FEATURES USED FOR THE PREDICTION OF THE DETECTABILITIES.

IV. PREDICTING DETECTABILITY

In order to provide useful information in a driver assistance context, we have to prove that a purposeful mapping can be learned which is able to reliably and robustly predict the detectabilitiy of pedestrians. Our dataset contains a total of 852 samples of pedestrians with associated detectabilities. We randomly split this dataset into 600 samples for training and the remaining 252 samples were set aside for testing and estimating the validation performance. The mapping can be realized in terms of a regression function. This is a difficult task since the training data obtained during the first experiment is noisy. Even though the subset of images from the StreetScenes was selected by hand, the images and annotations are not perfect. The manual imprecision of the participants can lead to false detections or misses and in situations where a group of pedestrians is close together, the response clicks of the participants can be ambiguous (see Figure 3 (right)). To counteract the noise in the data, we needed to train a regression function with the help of a machine learning algorithm that operates on a set of robust features. We chose Support Vector Regression (SVR, *c.f.* [14], [33]) with a Radial Basis Function (RBF) kernel as a regressor as it has proven to produce state-of-the-art results even on small and noisy datasets. We employed the freely available LIBSVM [11] implementation to train and test our SVRs. We extracted a large battery of features from the dataset (Table II shows a list of all features).

Features encoding the symmetry of the pedestrians such as *mfThres*, *mfCount*, *mfMaxScale* and *mfMeanScale* are of special interest. As observed by *e.g.* [4] and [9], shape-centered features that encode the symmetry can be very powerful for pedestrian detection and tracking since pedestrians are highly symmetric shapes. Therefore, we expect pedestrians that do not possess a symmetric shape to be less detectable by humans. Furthermore, several features depend on the difference between the pedestrian and its context (for a critical discussion of context for object detection see Wolf *et al.* [36]). We define the context of a pedestrian as a box three times the width and double the height of the pedestrian, located around the center of the pedestrian (see Figure 5).

The total dimensionality of all combined features is 378. We normalized the feature vectors to ensure that each dimension has a mean of zero and a standard deviation of one (variance normalization) in order to guarantee that no single dimension will dominate the ensuing distance calculations in feature space. Without normalization, distance judgments between two feature vectors could be dominated by one dimension, whose variance could be, for example, a couple of magnitudes larger than the rest. As the number of feature dimensions is high when compared to the number of training samples we would run into the so-called *curse of dimensionality* (see [3], [5]). This observation states that

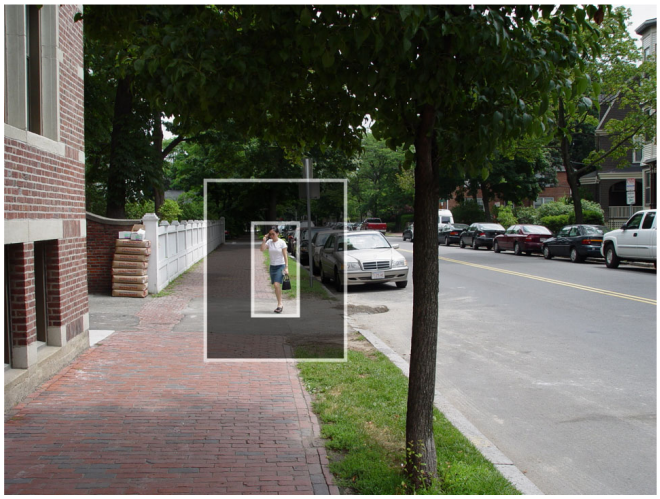

Fig. 5. The context of a pedestrian (gray) is the box twice the height of the pedestrian and three times its width around the center of the pedestrian.

as the dimensionality of a problem increases, the number of samples to evenly sample the space grows exponentially. Furthermore, we can assume that there is a certain amount of redundancy between dimensions (especially between different color histograms). Preprocessing with a Principle Component Analysis (PCA) did not improve the performance, most likely due to the high levels of noise in some of the features. Consequently, we opted for a feature selection technique to compute an optimal subset of features for the regression task (for more information and references on feature selection techniques confer to, *e.g.* Huan *et al.* [20]). As the number of possible feature-combinations grows exponentially with the number of features, a search for the optimum by enumerating all combinations is infeasible. Therefore, we used a straight-forward heuristic to simultaneously select features and optimize the free parameters of the SVR (the regularization parameter C, the slack variable ξ and the σ of the RBF Kernel). We randomly initialized all parameters and the feature selection, trained the SVR on the training set and used the cross-validation error as a measure of the performance. We trained a large number of machines (in the order of 20.000) in this manner while continuously keeping the best 100 SVRs. We then narrowed down the search space according to the parameter distribution of the remaining 100 SVRs and repeated the search with a more fine grained parameter sampling. We repeated this procedure till convergence. This method is closely related to *simulated annealing* and yields a result that is likely to be close to an optimum. Since this problem is easily parallelizable, as training one SVR is independent of the others, computation time is not an issue. Table III shows, which features were finally selected by our feature selection scheme. The dimensionality of the final descriptor is 68. Interestingly, the final descriptor contains two shape-centered features indicating the importance of symmetry for human pedestrian detection.

After the feature selection and the parameter optimization,

Name	Dimensionality
Area	1
PedCount	1
DiffStd	1
DiffHist	51
Dist2Center	1
Dist2Ped	1
DiffFix	1
PedMean	1
PixelPerClass	8
mfCount	1
mfMeanScale	1

TABLE III

THE RESULT OF OUR FEATURE SELECTION SCHEME. THE REDUCED FEATURE VECTOR CONTAINS ONLY 68 DIMENSIONS.

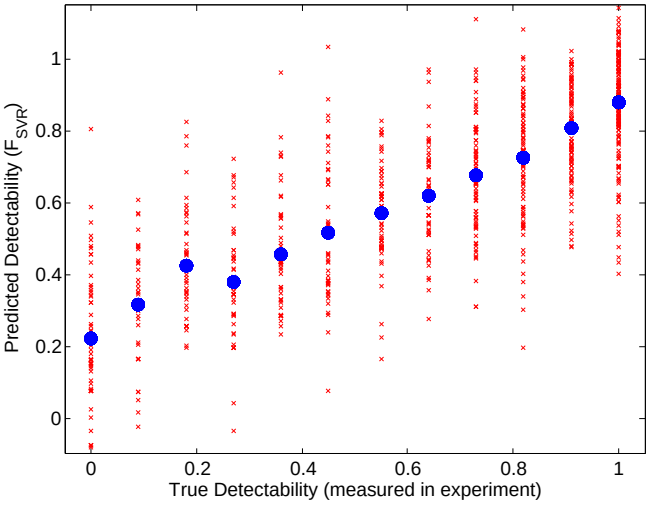

Fig. 6. Prediction performance of the trained regressor. The plot shows the predicted detectability versus the true detectability (as assessed by our experiment) for all images in the test dataset. Blue circles show the means for each level of ground-truth detectability. The means show a clear linear correlation.

we learned a model on the training data and predicted the detectabilities of all pedestrians in the test dataset:

$$F_{\text{SVR}} : \mathbf{X} \in \mathbb{R}^{68} \rightarrow \mathcal{D} \in [0, 1], \tag{2}$$

where F_{SVR} is the SVR we trained to predict the detectability $\mathcal{D}$ for a pedestrian with the feature vector $\mathbf{X}$. The results are plotted in Figure 6. The mean squared error of the prediction is 0.04 and the R^2 value for the correlation is 0.62. Due to the noise in the data, this might already be very close to the optimal performance for this dataset.

V. OPTIMIZING THE FOCUS OF ATTENTION

Finally, we want to demonstrate that being able to predict the detectability of a pedestrian has useful applications in a real world scenario.

A. Concept

Using the same experimental setup as before we now use our regression framework to maximize the overall detectability of all pedestrians in a scene. Specifically, we show that we

Fig. 7. Shown are pairs of images and corresponding heat maps of the predicted detectability for all possible fixation cross positions. Colors indicate the mean of the predicted detectability for all pedestrians in the image. The top row shows that our approach picks up on the large pedestrian in the foreground that is easily perceived and shifts the focus of attention to the harder to find pedestrians in the background.

Fig. 8. Examples of the two different methods of predicting the position of the fixation cross. The red dot indicates the center between all labeled pedestrians in the image while the blue dot is the optimal fixation point according to our regressor. In the two cases here, the regressor has estimated that one person will be particularly hard to detect and that the focus should be shifted closer to that pedestrian.

are able to predict an optimal location of the fixation cross. The optimal fixation cross location is the position where the probability that all pedestrians will be detected is maximal. In a driver assistance context this would be the location where, e.g., a head-up display would need to direct the attention of the driver in a sudden hazard situation. Four of our features (*Pos*, *Dist2Center*, *FixColor* and *DiffFix*, *c.f.* Table III) depend on the fixation point. By evaluating our regressor at all possible fixation positions (all possible image scene positions) and for all pedestrians detected in a scene we can determine the optimal fixation cross where the probability for the correct reporting, and thus recognition, of all pedestrians is maximized. Figure 7 (right) shows examples of the mean detectability for all pedestrians in the image as a function of the position of the fixation cross.

Without knowledge about the detectabilities a driver assistance system would have to assume that all pedestrians in the scene are equally difficult to spot and it would consequently have to guide the attention of the driver to the center of gravity of all pedestrians. This scheme yields different fixation points than the fixation points that have been predicted by our scheme (see samples in Figure 8).

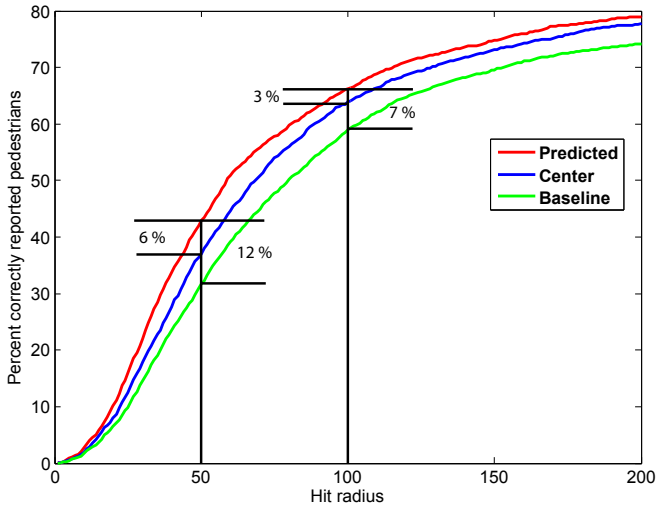

Fig. 9. Portion of pedestrians whose position was correctly reported and counted as a 'hit' as a function of the radius of the disc associated with the pedestrian. Our method based on the results of the regressor (red line) outperforms the methods that have no access to the detectability of the pedestrians (center of all pedestrians in the image (blue line) and random fixation cross position (green line)). The 6% absolute performance increase equals a relative improvement of 16% over the 'center' method.

B. Validation

Based on these two kinds of fixation cross prediction schemes (prediction using our regressor and the mean of the pedestrian positions), we set up a second experiment to evaluate whether our new method actually increases the overall pedestrian detectability. We selected 115 images from the database in which the two kinds of fixation points differed by at least 50 pixels. We repeated the experiment described in Section IV, but this time with a variable positioning of the fixation cross.

To augment the test data set and to better compare the two kinds of fixation locations, we used the mirrored versions of the images as well, and presented the two different fixation cross positions on the two versions of the same image. The pairing of fixation cross type, whether the image was mirrored or not and the presentation order was randomized for each participant. In addition, we added a 'baseline' condition in which the fixation cross position was unrelated to the image content. Random positioning would be an unfair baseline condition as this would include fixation crosses in the corners which is obviously suboptimal for the task. We took fixation cross positions from different images, predicted by one of the other two conditions, for our baseline condition. This allowed us to ensure similar spatial distributions of fixation crosses across all three conditions.

Each of the ten participants (mean age: 24.6 years, 5 males, 5 females) did 550 trials and completed a questionnaire afterwards. The head position of the participants was again fixated 65 centimeters away from the monitor using a chinrest resulting in a horizontal viewing angle of approximately 60°. Three of the subjects reported afterwards that some of the images were presented twice in mirrored conditions. Figure 9 shows the percentage of correctly re-

ported pedestrians across all trials as a function of hit radius (the radius of the ring in Figure 3 for all three kinds of fixation point locations (predicted, centered and random)). The Figure plots the portion of pedestrians that were correctly reported during the experiment as a function of hit-radius (distance from the pedestrian in which a click is counted as a detection of the pedestrian). Our regression based method outperformed the fixation cross positioning at the center between all pedestrians and the random baseline condition for all hit-radii. This proof-of-concept demonstrates that the ability to predict the detectabilities of pedestrians can be a useful source of information for adaptive HMI systems.

VI. SUMMARY & OUTLOOK

In this paper, we presented and evaluated the novel concept of estimating the detectability of pedestrians in natural images using machine vision. We trained a regressor to predict these detectabilities and were able to show that estimating the detectabilities can yield considerable improvements in the overall detection rate of pedestrians.

Future investigations could aim to evaluate the detectabilities of pedestrians in dynamic scenes. This would require either videos with high quality annotations or a virtual reality setup. A virtual reality setup would also allow closer control over the parameters that influence the detectability of pedestrians, the controlled introduction of distractors and state-dependent estimations (*e.g.* the current body pose of the pedestrian). This will allow us to evaluate our method in a broader range of settings.

REFERENCES

[1] A. Agarwal and B. Triggs. Monocular human motion capture with a mixture of regressors. In *Computer Vision and Pattern Recgonition*, pages 72–81. IEEE Computer Society, 2005.

[2] M. Andriluka, S. Roth, and B. Schiele. People-tracking-by-detection and people-detection-by-tracking. In *Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[3] R. Bellman. *Dynamic Programming*. Dover Publications, March 1957.

[4] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi. Shape-based pedestrian detection and localization. In *Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2003*, pages 328–333, Shangai, China, October 2003.

[5] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is "nearest neighbor" meaningful? In *International Conference on Database Theory*, pages 217–235, 1999.

[6] S. M. Bileschi. *Streetscenes: towards scene understanding in still images*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.

[7] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *Conference on Image and Video Retrieval*, pages 401–408. ACM, 2007.

[8] D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:433–436, 1997.

[9] T. Bücher, C. Curio, H. Edelbrunner, C. Igel, D. Kastrup, I. Leefken, G. Lorenz, A. Steinhage, and W. von Seelen. Image processing and behaviour planning for intelligent vehicles. *IEEE Transactions on Industrial Electronics*, pages 62–75, 2003.

[10] R. H. S. Carpenter. *Movements of the eyes*. Pion, London, 1977.

[11] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001.

[12] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *International Conference on Computer Vision & Pattern Recognition*, pages 886–893, 2005.

[13] A. Doshi and M. Trivedi. Attention estimation by simultaneous observation of viewer and view. In *Computer Vision and Pattern Recognition Workshops*, pages 21 – 27, 2010.

[14] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support vector regression machines. In *Advances in Neural Information Processing Systems*, pages 155–161. MIT Press, 1997.

[15] D. Engel and C. Curio. Scale-invariant medial features based on gradient vector flow fields. In *International Conference on Pattern Recognition*, 2008.

[16] M. Enzweiler and D. M. Gavrila. Monocular Pedestrian Detection: Survey and Experiments. *Pattern Analysis and Machine Intelligence*, 31(12):2179–2195, 2008.

[17] L. Fletcher, G. Loy, N. Barnes, and A. Zelinsky. Correlating driver gaze with the road scene for driver assistance systems. *Robotics and Autonomous Systems*, 52(1):71 – 84, 2005.

[18] C. Fulgenzi, A. Spalanzani, and C. Laugier. Probabilistic motion planning among moving obstacles following typical motion patterns. In *International conference on Intelligent Robots and Systems*, pages 4027–4033. IEEE Press, 2009.

[19] J. Haushofer, P. H. Schiller, G. Kendall, W. M. Slocum, and A. S. Tolias. Express saccades: the conditions under which they are realized and the brain structures involved. *Journal of Vision*, 2(7):174–174, 2002.

[20] L. Huan and H. Motoda. *Feature Selection for Knowledge Discovery and Data Mining*. Springer, 1998.

[21] H. Kirchner and S. J. Thorpe. Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11):1762 – 1776, 2006.

[22] M. Kleiner, D. Brainard, and D. Pelli. What's new in psychtoolbox-3? In *European Conference on Visual Perception*, 2007.

[23] T. Lan, Y. Wang, W. Yang, and G. Mori. Beyond actions: Discriminative models for contextual group activities. In *Advances in Neural Information Processing Systems (NIPS)*, 2010.

[24] C. Laugier, S. Petti, D. A. Vasquez Govea, M. Yguel, T. Fraichard, and O. Aycard. Steps towards safe navigation in open and dynamic environments. In *Conference on Robotics and Automation*, 2005.

[25] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *International Journal of Computer Vision*, 77(1-3):259–289, 2008.

[26] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In *Computer Vision and Pattern Recognition*, pages 878–885, 2005.

[27] K. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. In *European Conference on Computer Vision*, volume 1, pages 69–81, 2004.

[28] R. Okada and S. Soatto. Relevant feature selection for human pose estimation and localization in cluttered images. In *European Conference for Computer Vision*, pages 434–445, 2008.

[29] C. Papageorgiou, T. Evgeniou, and T. Poggio. A trainable pedestrian detection system. In *Proceedings of Intelligent Vehicles*, pages 241–246, 1998.

[30] S. Pinneli and D. M. Chandler. A Bayesian approach to predicting the perceived interest of objects. *International Conference on Image Processing*, pages 2584–2587, 2008.

[31] L. Pomarjanschi, M. Dorr, C. Rasche, and E. Barth. Safer driving with gaze guidance. In *Proceedings of Bionetics*, 2010. in press.

[32] Bundesamt für Straßenwesen. Verkehrs- und Unfalldaten. http://www.bast.de, 2010.

[33] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[34] L. Sigal, S. Bhatia, S. Roth, M. J. Black, and M. Isard. Tracking loose-limbed people. *Computer Vision and Patter Recognition*, 01:421–428, 2004.

[35] M. Spain and P. Perona. Measuring and Predicting Object Importance. *International Journal of Computer Vision*, 91(1):59–76, Aug. 2010.

[36] L. Wolf and S. Bileschi. A critical view of context. *International Journal of Computer Vision*, 69(2):251–261, 2006.