Monocular Pedestrian Recognition Using Motion Parallax

M. Enzweiler[1], P. Kanter[2] and D. M. Gavrila[2,3]

[1] Image & Pattern Analysis Group,
Department of Mathematics and Computer Science,
University of Heidelberg, 69115 Heidelberg, Germany

[2] Environment Perception, Group Research, Daimler AG,
Wilhelm-Runge-St. 11, 89081 Ulm, Germany

[3] Intelligent Systems Lab, Faculty of Science, University of Amsterdam,
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

{uni-heidelberg.enzweiler,dariu.gavrila}@daimler.com

Abstract— **This paper presents a novel focus-of-attention strategy for monocular pedestrian recognition. It uses Bayes' rule to estimate the posterior for the presence of a pedestrian in a certain (rectangular) image region, based on motion parallax features. This posterior is used as a parameter to control the amount of regions of interest (ROIs) that is passed to subsequent verification stages. For the latter, we use a state-of-the-art pedestrian recognition scheme which consists of multiple modules in a cascade architecture. We obtain optimized settings for the control parameters of the combined cascade system by a sequential ROC convex hull technique.**

Experiments are conducted on image data captured from a moving vehicle in an urban environment. We demonstrate that the proposed focus-of-attention strategy reduces the false positives of an otherwise identical monocular pedestrian recognition system by a factor of two, at equal detection rates. The overall system maintains processing rates close to real-time.

I. Introduction

Pedestrian protection is an important problem in the domain of intelligent vehicles. Large variations in pedestrian appearance (e.g. clothing, pose) and environmental conditions (e.g. lighting, background) make this problem particularly challenging. A typical approach starts by identifying regions of interest (ROIs) in the image using a computationally efficient method (e.g. motion detection, obstacle detection) and thereafter moves on to a more expensive pattern classification step. In this paper, we focus on the generation of regions of interest.

Stereo vision has been proposed as a powerful technique to obtain initial ROIs [9], [19]. However, the use of stereo vision also involves some drawbacks, such as the cost of an additional camera and adverse effects on design. The objective of the paper is to see to what degree one can rely on monocular processing. In particular, this paper aims to improve performance by an early attentive stage based on motion parallax features.

The general idea of early focus-of-attention is independent of the actual pedestrian recognition system used. We integrate the proposed ROI generation technique with real-time (monocular) shape-texture based pedestrian detection and tracking, as proposed in [9]: Shape-based detection is achieved by efficient matching of an exemplar-based shape hierarchy to the generated ROIs. Shape matches are verified by a texture-based pedestrian classifier, a neural network operating on local adaptive receptive fields. Temporal integration is provided by an α-β tracker. See Figure 1.

II. Related Work

Recently, [8] presented a survey on pedestrian protection systems in the intelligent vehicles domain, focusing on active and passive pedestrian safety, as well as risk assessment strategies using multiple vision and non-vision sensors. In this paper, we concentrate on vision-based monocular pedestrian recognition in the visible spectrum, where we limit ourselves to the case of (almost) fully visible pedestrians in an upright pose.

A sizeable body of literature exists on the recognition of pedestrians in images. Roughly, approaches can be categorized in terms of the utilized feature sets. Some approaches employ spatial features, such as shape or local texture [6], [9], [10], [11], [12], [15], [17], [18], designed to detect static pedestrians ([11] provides a large public dataset for benchmarking). Temporal information is usually incorporated in terms of an independent tracking algorithm, e.g. [3], [9].

Other techniques involve purely temporal cues to model characteristical motion patterns resulting from human gait, in terms of (relative) deviations in the optical flow field [7], [13], [14], or hand-crafted human motion models [5]. Such approaches usually fail to detect static pedestrians, due to the lack of temporal cues on static objects.

The most popular strategy to exploit characteristical human motion patterns has involved enriching spatial representations with temporal information [6], [15], [17], [18].
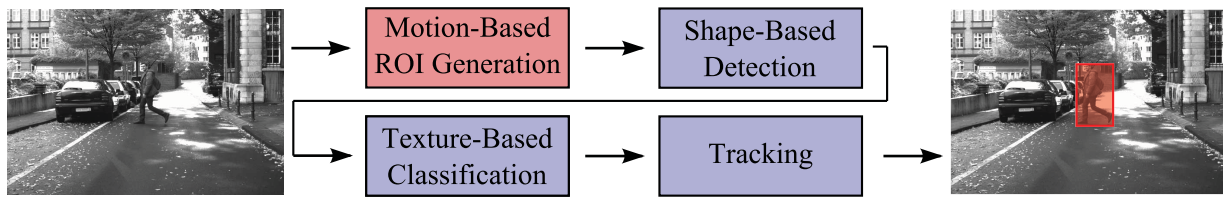
Fig. 1. Overview of the integrated pedestrian recognition system comprising motion-based ROI generation, shape-based detection, texture-based classification and tracking. We focus on the motion-based ROI generation module (red), but evaluate the whole integrated system.

Feature dictionaries based on non-adaptive Haar wavelets [17] or local shape filters [15] have been extended to the spatio-temporal domain by additionally incorporating temporal intensity differences. However, using representations of absolute motion requires targets moving in front of a largely static camera and background.

To overcome this drawback, adaptive spatio-temporal features using relative codings of motion features have been proposed [6], [18], as follows. First, adaptive spatial intensity-based local receptive field features have been generalized to spatio-temporal receptive fields [18], to infer both salient appearance and characteristical motion patterns of humans from training data. Second, histograms of oriented gradients (HOG) have been coupled with a relative motion encoding scheme based on histograms of differential optical flow [6]. Several researchers have compared the performance of otherwise identical spatial and spatio-temporal features and reported superior performance of the spatio-temporal feature sets at the drawback of requiring temporally aligned training samples [6], [17].

In light of previous research, we propose a motion-based attention mechanism to improve the performance of a state-of-the-art monocular shape-texture-based pedestrian recognition system. This attentive strategy utilizes a learned probabilistic model of motion-based features, which are particularly attuned to pedestrians. The features involve mean horizontal velocity and density of local parallax flow. The application of parallax flow allows to focus on static non-planar or moving objects, while at the same time compensating camera ego-motion. Further, this representation seamlessly extends to the detection of static pedestrians, unlike previous approaches which require target motion [5], [7], [13], [14].

We do not consider a mixture of spatial and temporal features within a single classifier, as in [6], [15], [17], [18]. Instead, we employ a cascade structure with complementary cues for each module, as shown in Figure 1, to successively narrow down the search space, see [9]. The proposed motion-model is utilized as ROI generation module for subsequent shape- and texture-based pedestrian classification, based on a sound Bayesian assessment of posterior probabilities for each ROI. Parameters of the integrated multi-cue system are optimized with regard to robustness and efficiency for maximum real-time performance, by employing sequential ROC optimization [9].

III. MOTION-BASED PEDESTRIAN MODEL

The proposed probabilistic motion-based pedestrian model is based on optical flow features [2], induced by our moving camera. Rather than directly utilizing the observed intrinsic flow field, camera ego-motion is canceled out (estimated from inertial sensors) by computing the parallax flow field [1]. Parallax flow is the difference between the intrinsic optical flow field and estimated ground-plane flow. Residual parallax flow vectors are then induced by both static non-planar and moving objects. See Figure 2.

Given a sparse parallax flow field $P(\mathbf{x}) = \mathbf{v_x}$, let $\mathbf{v_x} = (u_\mathbf{x}, v_\mathbf{x})^\top$ denote horizontal and vertical components of parallax flow at location $\mathbf{x} = (x, y)^\top$. Further, we introduce a function $S_P(\mathbf{x})$, assessing sparseness of a given parallax flow field P at location $\mathbf{x}$, with $S_P(\mathbf{x}) = 1$, if $\mathbf{v_x}$ exists, and $S_P(\mathbf{x}) = 0$ otherwise. For a particular (rectangular) region
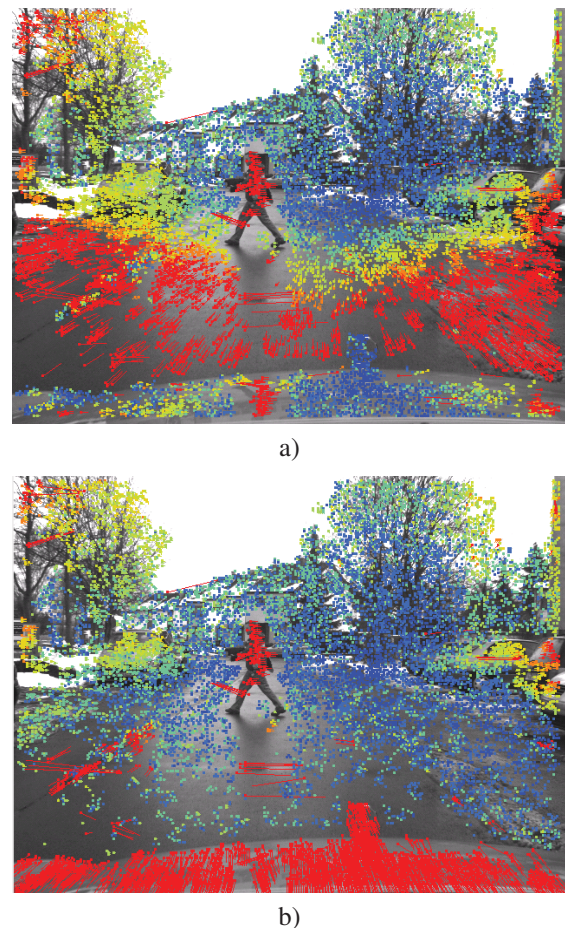

Fig. 2. Visualization of considered optical flow fields. Warmer colors encode longer flow vectors. a) observed intrinsic flow, b) ego-motion corrected parallax flow focusing on static non-planar and moving objects
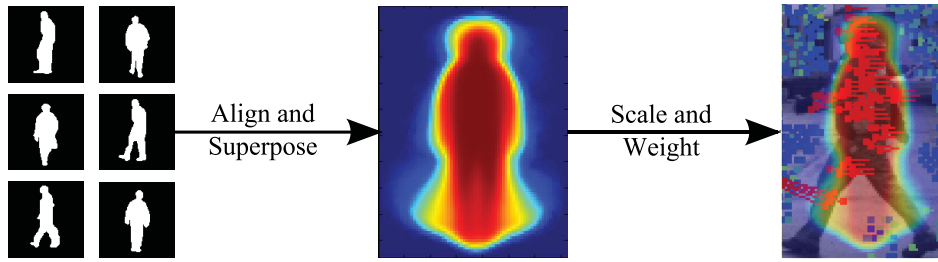
793

Fig. 3. Overview of the probabilistic weighting scheme to adapt the motion-based flow-features to the pedestrian class. A two-dimensional probability mass function (center) is learned from binary pedestrian foreground masks (left) and utilized to weight feature computation (right). Warmer colors encode higher probability.

of interest R, our aim is to estimate its posterior probability, $\mathbf{P}(\omega_0|R)$, with respect to the pedestrian class ω_0. To that extent, we represent R in terms of a feature set φ_R based on parallax flow and follow a Bayesian approach:

$$\mathbf{P}(\omega_0|R) = \mathbf{P}(\omega_0|\varphi_R) = \frac{\mathbf{p}(\varphi_R|\omega_0)\mathbf{P}(\omega_0)}{\sum_{i=0}^{1}\mathbf{p}(\varphi_R|\omega_i)\mathbf{P}(\omega_i)} \quad (1)$$

Priors for the pedestrian ω_0 and non-pedestrian class ω_1, $\mathbf{P}(\omega_0)$ and $\mathbf{P}(\omega_1)$, are assumed uniform. In the following, details on feature selection and the estimation of the likelihoods $\mathbf{p}(\varphi_R|\omega_i)$ are given.

Selecting appropriate motion-based features φ_R involves a trade-off between generality and specificity. On the one hand, features should be general with respect to arbitrary pedestrian appearance and motion, while on the other hand, powerful distinction between pedestrians and background is desired. In view of utilizing the proposed model as a focus-of-attention strategy for subsequent classification modules, our main concern at this point is generality. Hence, the proposed motion-based features involve rather generic measures to not reject potential pedestrian candidate regions too early in the processing cascade. We consider mean horizontal velocity V_R and density D_R within a local ROI R. To enhance specificity, features are particularly attuned to pedestrians in terms of both a probabilistic weighting scheme and statistical combination.

Pedestrian motion typically involves a characteristic velocity range, as opposed to other moving objects in urban traffic. Given that pedestrian motion is predominantly horizontal, we restrict ourselves to the horizontal velocity component u of parallax flow vectors. At the same time, errors induced by excessive pitch-movement of the camera are alleviated. Let $\mathbf{P}_w(\mathbf{x}_r|\omega_0)$ denote a location-specific probabilistic weighting scheme for $\mathbf{x}_r \in R$ which is employed to adapt the proposed features to the pedestrian class (the definition of $\mathbf{P}_w(\mathbf{x}_r|\omega_0)$ is given below). Then, V_R involves the weighted mean of horizontal flow components, $u_{\mathbf{x}_r}$, at locations $\mathbf{x}_r \in R$:

$$V_R = \frac{\sum_{\mathbf{x}_r \in R} |u_{\mathbf{x}_r}|S_P(\mathbf{x}_r)\mathbf{P}_w(\mathbf{x}_r|\omega_0)}{\sum_{\mathbf{x}_r \in R} S_P(\mathbf{x}_r)\mathbf{P}_w(\mathbf{x}_r|\omega_0)} \quad (2)$$

As a second feature, we propose weighted mean flow density D_R within R, again utilizing $\mathbf{P}_w(\mathbf{x}_r|\omega_0)$ as weighting scheme:

$$D_R = \frac{\sum_{\mathbf{x}_r \in R} S_P(\mathbf{x}_r)\mathbf{P}_w(\mathbf{x}_r|\omega_0)}{\sum_{\mathbf{x}_r \in R} \mathbf{P}_w(\mathbf{x}_r|\omega_0)} \quad (3)$$

Local flow density in regions corresponding to pedestrians is expected to be rather sparse. The highly articulated and non-rigid pedestrian motion, combined with continuously appearing and disappearing background, as well as self-occlusions, negatively affects the computation of pixel-wise correspondences over time. Hence, the local density measure aims to distinct pedestrians from largely rigid objects, where recovered flow estimates are taken to be more dense. See Figure 2.

$\mathbf{P}_w(\mathbf{x}_r|\omega_0)$ denotes a two-dimensional probability mass function, representing the likelihood that a given location $\mathbf{x}_r \in R$ corresponds to a pedestrian. To estimate $\mathbf{P}_w(\mathbf{x}_r|\omega_0)$, the superposition of a set of n aligned binary pedestrian foreground masks, $f_s(\mathbf{x}_r)$, as defined by manually labeled pedestrian contours, is utilized, see Figure 3:

$$\mathbf{P}_w(\mathbf{x}_r|\omega_0) \sim \sum_{s=1}^{n} f_s(\mathbf{x}_r), \quad (4)$$

with $\mathbf{P}_w$ spatially scaled to the dimensions of the ROI R and normalized such that

$$\sum_{\mathbf{x}_r \in R} \mathbf{P}_w(\mathbf{x}_r|\omega_0) = 1. \quad (5)$$

To further increase the discriminative power and robustness of the proposed features, we consider statistically combining V_R and D_R into a multidimensional feature $\varphi_R = V_R \wedge D_R$. Under the assumption of independence of V_R and D_R, the likelihood functions in Eq. (1) can be decomposed into:

$$\mathbf{p}(\varphi_R|\omega_i) = \mathbf{p}(V_R \wedge D_R|\omega_i) = \mathbf{p}(V_R|\omega_i)\mathbf{p}(D_R|\omega_i) \quad (6)$$

Approximations of $\mathbf{p}(V_R|\omega_i)$ and $\mathbf{p}(D_R|\omega_i)$ are obtained via histogramming of training data with regard to the proposed features. In case of pedestrians, we utilize manually labeled bounding boxes, whereas non-pedestrian labels are randomly extracted from parallax flow fields of non-pedestrian images using ground-plane constraints.
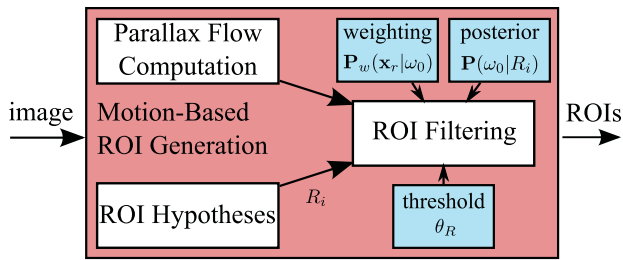
794

Fig. 4. Overview of motion-based ROI generation (cf. Figure 1): parallax flow computation, generation of ROI hypotheses and filtering of ROI hypotheses.

IV. SYSTEM INTEGRATION

The probabilistic motion-based pedestrian model, as presented in Section III, is utilized as ROI generation module, see Figure 1. This module involves three components: optical (parallax) flow computation, generation of ROI hypotheses and filtering of ROI hypotheses. The filtered ROIs define initial search areas for the subsequent detection module. An overview of these subcomponents is given in Figure 4.

We consider the proposed flow-based features (see Section III) as independent from the actual algorithm to compute optical flow, see [2]. In our experiments, an efficient technique involving correspondences based on census transform signatures is utilized [16]. This allows for real-time flow computation (25 Hz). Parallax flow is obtained by canceling out camera ego-motion, estimated from inertial sensors [1].

Initial object location hypotheses, R_i, are generated using a sliding window technique, where detector windows at various scales and locations are shifted over the image. Here, application-specific constraints, such as flat-world assumption, people standing on the ground or prior knowledge about the dimensions of target objects, are incorporated, see Figure 5a. Each pedestrian candidate region R_i is represented in terms of features V_{R_i} and D_{R_i}, followed by the estimation of posterior probability with respect to the pedestrian class, $\mathbf{P}(\omega_0|R_i)$, see Eq. (1). A threshold θ_R governs the amount of ROIs which are committed to the subsequent module: Only ROIs with $\mathbf{P}(\omega_0|R_i) > \theta_R$, as shown in Figure 5b, trigger the evaluation of the next cascade module. Others are rejected immediately.

Pedestrian recognition proceeds with shape-based detection (see Figure 5c), involving coarse-to-fine matching of an exemplar-based shape hierarchy to the image data at hand [9]. Positional initialization is given by the output ROIs of the motion-based attention stage. The shape hierarchy is constructed off-line in an automatic fashion from manually annotated shape labels. On-line matching involves traversing the shape hierarchy with the Chamfer distance between a shape template and an image sub-window as smooth and robust similarity measure [4]. Image locations, where the similarity between shape and image is above a user-specified threshold, are considered detections. A single distance threshold applies for each level of the hierarchy. Additional parameters govern the edge density on which the
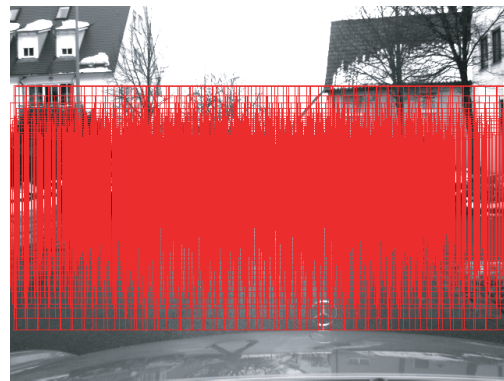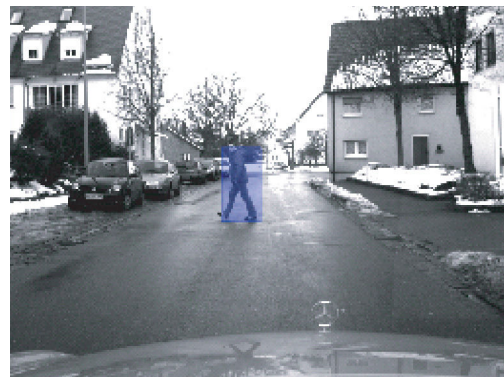

a)
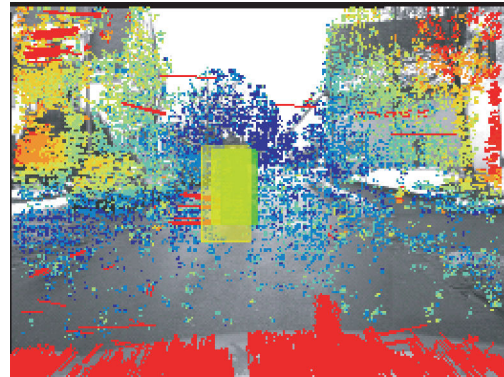
b)
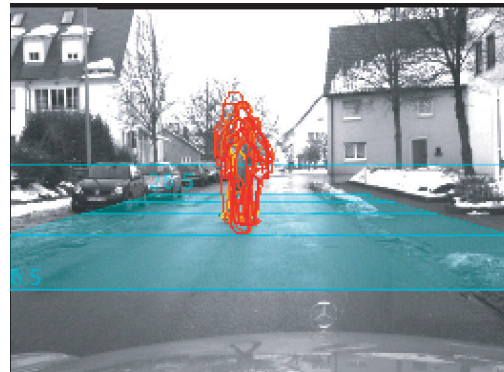
c)

d)

Fig. 5. Results of the integrated pedestrian recognition system. a) initial ROI hypotheses, b) ROIs filtered using the proposed motion-based focus-of-attention strategy, c) results of shape-based detection, d) system output after texture-based classification and tracking

795

	dataset 1	dataset 2	dataset 3
# of images	2970	2000	2702
# of pedestrian trajectories	45	52	38
pedestrian labels	1 606	1 302	1 625
non-pedestrian labels	910 369	856 256	832 639

TABLE I

STATISTICS OF CONSIDERED DATA SETS.

underlying distance map is based.

Detections of the shape matching step are verified by a texture-based pattern classifier. We employ a multi layer feed-forward neural network operating on local adaptive receptive field features [9], [18]. Finally, temporal integration of detection results is employed to overcome gaps in detection and suppress spurious false positives. A 2D bounding box tracker is utilized, with an object state model involving bounding box position and extent [9]. State parameters are estimated using an $\alpha - \beta$ tracker, see Figure 5d.

V. EXPERIMENTS

The proposed motion-based attention strategy was tested in experiments on pedestrian recognition from a moving vehicle. Datasets were acquired in daylight conditions in urban traffic and depict non-occluded pedestrians in front of a changing background. Pedestrian labels were manually extracted, whereas non-pedestrian labels were obtained randomly from non-pedestrian images using a sliding window technique in conjunction with ground-plane constraints. See Table I for the datasets used. In all experiments, we perform threefold cross-validation: Two datasets are utilized at a time to learn the probabilistic model of motion-features and to optimize parameters, respectively. Performance is evaluated on the remaining dataset.

In a first experiment, we evaluate the ability of the motion module (Section III) to select the relevant ROIs using the proposed features. In particular, we consider both mean horizontal velocity and density as single features, $\varphi_R = V_R$ and $\varphi_R = D_R$, as well as the statistically combined multi-dimensional feature $\varphi_R = V_R \wedge D_R$, see Eq. (6). To evaluate the inherent quality of flow features, the manually labeled pedestrians and corresponding non-pedestrians (see Table I) are directly employed as training and test sets. We thus consider classification performance of the motion module. A test sample R is classified as pedestrian ω_0, if the associated posterior probability $\mathbf{P}(\omega_0|\varphi_R) > \theta_R$, see Eq. (1). Figure 6a shows the performances of different features sets φ_R, in terms of mean ROC curves of all three cross-validation runs, with threshold θ_R varied along the curves. It is observed, that the mean horizontal velocity feature V_R is superior to the density feature D_R. Further performance boost is achieved by statistically combining both features to a robust multi-dimensional feature $\varphi_R = V_R \wedge D_R$, see Eq. (6).

We now turn our attention to the problem of pedestrian recognition using test sequences consisting of images with a size of 640×480 pixels, see Table I. The proposed integrated system using motion-based ROI generation (see Figure 1) is compared to an otherwise identical monocular recogni-
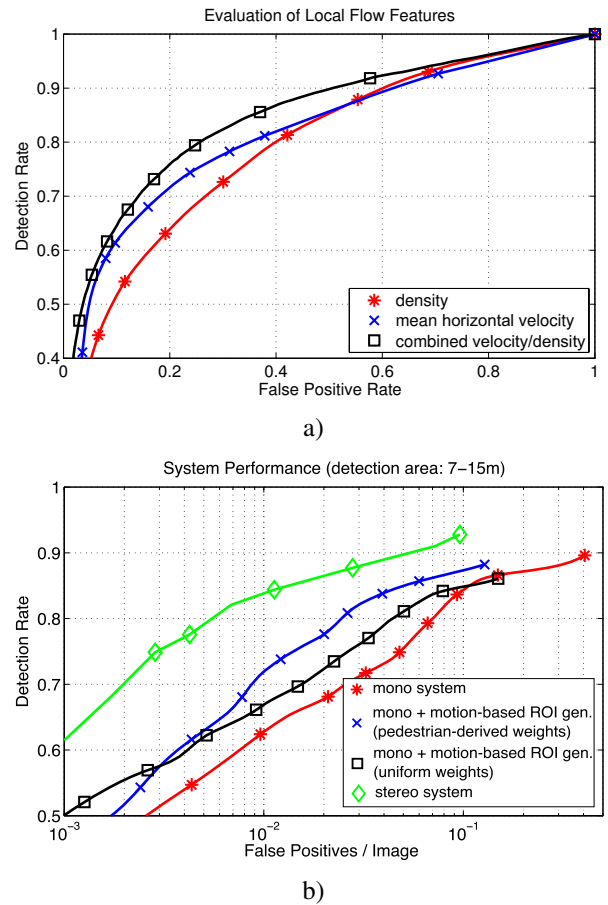

a)

b)

Fig. 6. Mean ROC performance of three cross-validation runs for a) evaluation of local flow features, b) different variants of pedestrian recognition systems. Note: scale is logarithmic

tion system using the ground-plane as sole ROI constraint for shape-matching. Further, we compare to a stereo-based pedestrian recognition system which uses depth information for ROI generation, see [9].

The motion-based ROI generation module utilizes the combined velocity / density feature (the best performing variant in Figure 6a) in conjunction with two different feature weighting strategies: First, we employ the proposed probabilistic weighting scheme derived from pedestrian foreground masks, see Figure 3, with the weight for each location $\mathbf{x}_r$ within an ROI R, $\mathbf{P}_w(\mathbf{x}_r|\omega_0)$, determined using Eqs. (4) and (5). Further, we consider equal weights, with $\mathbf{P}_w(\mathbf{x}_r|\omega_0)$ defined as uniform distribution.

To balance efficiency and robustness for maximum performance of all considered systems, significant parameters of each detection module (see Figure 1) have been optimized using sequential ROC optimization [9]. Parameters subject to optimization include the posterior threshold θ_R for motion-based ROI generation, edge and distance thresholds for shape-based detection, output threshold for texture-based classification, as well as track start and termination criteria for tracking. This technique avoids ad-hoc parameter tuning and provides tight module integration.

Evaluation criteria and application-specific requirements

796

	mono	mono + motion-based ROI gen. (pedestrian-derived weights)	mono + motion-based ROI gen. (uniform weights + integral images)	stereo
images per second	8.1 Hz	7.2 Hz	15 Hz	15.5 Hz
processing time per image	123 ms	138 ms	67.1 ms	64.5 ms

TABLE II

PROCESSING SPEED OF CONSIDERED PEDESTRIAN RECOGNITION SYSTEMS.

for pedestrian recognition are specified in 3D. In particular, the sensor coverage area is defined in relation to the vehicle as $7m-15m$ in longitudinal and $\pm 3m$ in lateral direction. Only fully-visible ground-truth pedestrians within the sensor coverage area are considered required, others are regarded as optional, in the sense that systems are not rewarded/penalized for correct/false/missing detections.

Given a system alarm and a ground-truth event, we enforce a maximum positional deviation in 3D to count the alarm as match, where both 2D ground-truth and 2D detections are back-projected into 3D using known camera geometry and ground-plane constraints. Localization tolerances are defined as percentage of distance for lateral (X) and longitudinal (Z) direction with respect to the vehicle. Here, we consider tolerances of $X = 10\%$ and $Z = 30\%$ with a larger tolerance in longitudinal direction to account for non-flat road surface and vehicle pitch in case of back-projection of (monocular) ground-truth and detections into 3D.

Performance is given in terms of mean ROC curves over three cross-validation runs, depicting system performance (detection rate vs. false positives per frame) after the final module (tracking) for each system under consideration. From Figure 6b it is observed, that the presented attentive strategy involving motion-based ROI generation improves performance of an otherwise identical monocular pedestrian recognition system, even if uniform feature-weights are used (black curve vs. red curve). Additional performance gain is achieved by increasing feature-specificity in terms of the proposed probabilistic weighting scheme, derived from pedestrian foreground masks (blue curve). Compared to the monocular system without any attention mechanism, false positives are significantly reduced by a factor of two, at equal detection rates (blue curve vs. red curve). The system variant utilizing stereo vision to obtain initial ROIs, outperforms all other monocular approaches by up to an order of magnitude (green curve).

Processing time has been evaluated using implementations in C/C++ on a state-of-the-art 2.66 GHz Intel PC, see Table II. Compared to the regular mono pedestrian recognition system, the proposed motion-based attention strategy using non-uniform weighting yields a significant boost in detection performance, paid for with only a minor increase in processing time (7.2 Hz vs. 8.1 Hz). Using a uniform feature-weighting strategy results in a significant reduction of computational costs (15 Hz vs. 7.2 Hz) at the cost of a decrease in detection performance, since the motion-based features V_R and D_R are less specific to the pedestrian class. This cut of computational costs is due to the fact that uniform weighting allows to exploit integral images, as proposed by [17], to compute the motion-based features. The approach

employing stereo vision exhibits both the best recognition performance and the lowest processing costs (15.5 Hz).

VI. CONCLUSION

This paper presented a novel attentive strategy for monocular pedestrian recognition involving a model of motion-based features learned from ego-motion corrected optical flow. Features are particularly attuned to the pedestrian class and modeled in a probabilistic fashion. In experiments on datasets captured from a moving vehicle in urban traffic, we obtained the result that pedestrian recognition performance is substantially enhanced by the proposed motion-based attention concept; false positives were reduced by a factor of two. Future work involves the incorporation of other sensor architectures and the Bayesian integration of individual detection modules.

REFERENCES

[1] D. Baehring et al. Detection of close cut-in and overtaking vehicles for driver assistance based on planar parallax. *IEEE IV Symp.*, pages 290–295, 2005.
[2] S. S. Beauchemin and J. L. Barron. The computation of optical flow. *ACM Comp. Surv.*, 27(3):433–467, 1995.
[3] M. Bertozzi et al. Pedestrian localization and tracking system with Kalman filtering. *IEEE IV Symp.*, pages 584–589, 2004.
[4] G. Borgefors. Distance transformations in digital images. *CVGIP*, 34(3):344–371, 1986.
[5] C. Curio et al. Walking pedestrian recognition. *IEEE ITS*, 1(3):155–163, 2000.
[6] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In *Proc. ECCV*, pages 428–441, 2006.
[7] B. Fardi et al. Motion-based pedestrian recognition from a moving vehicle. *IEEE IV Symp.*, pages 219–224, 2006.
[8] T. Gandhi and M. M. Trivedi. Pedestrian protection systems: Issues, survey, and challenges. *IEEE ITS*, 8(3):413–430, 2007.
[9] D. M. Gavrila and S. Munder. Multi-cue pedestrian detection and tracking from a moving vehicle. *IJCV*, 73(1):41–59, 2007.
[10] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In *Proc. CVPR*, pages 878–885, 2005.
[11] S. Munder and D. M. Gavrila. An experimental study on pedestrian classification. *IEEE PAMI*, 28(11):1863–1868, 2006.
[12] C. Papageorgiou and T. Poggio. A trainable system for object detection. *IJCV*, 38:15–33, 2000.
[13] R. Polana and R. Nelson. Low-level recognition of human motion. In *IEEE Work. on Mot. of Non-Rigid and Artic. Obj.*, pages 77–92, 1994.
[14] A. Shashua, Y. Gdalyahu, and G. Hayon. Pedestrian detection for driving assistance systems: Single-frame classification and system level performance. *IEEE IV Symp.*, pages 1–6, 2004.
[15] H. Sidenbladh and M. J. Black. Learning the statistics of people in images and video. *IJCV*, 54(1/2/3):183–209, 2003.
[16] F. Stein. Efficient computation of optical flow using the census transform. In *Proc. DAGM*, pages 79–86, 2004.
[17] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *IJCV*, 63(2):153–161, 2005.
[18] C. Wöhler and J. Anlauf. An adaptable time-delay neural-network algorithm for image sequence analysis. *IEEE Transactions on Neural Networks*, 10(6):1531–1536, 1999.
[19] L. Zhao and C. Thorpe. Stereo and neural network-based pedestrian detection. *IEEE ITS*, 1(3):148–154, 2000.