

Head Detection and Orientation Estimation for Pedestrian Safety

Eike Rehder¹, Horst Kloeden² and Christoph Stiller¹

Abstract—Head detection and orientation estimation are a vital component in the intention recognition of pedestrians. In this paper we propose a novel framework to detect highly occluded pedestrians and estimate their head orientation. Detection is performed for pedestrian’s heads only. For this we employ a part-based classifier with HOG/SVM combinations. Head orientations are estimated using discrete orientation classifiers and LBP features. Results are improved by leveraging orientation estimation for head and torso as well as motion information. The orientation estimation is integrated over time using a Hidden Markov Model. From the discrete model we obtain a continuous head orientation. We evaluate our approach on image sequences with ground truth orientation measurements. To our best knowledge we outperform state of the art results.

I. INTRODUCTION

In the past twenty years the number of people injured in traffic accidents has decreased drastically. However the number of injured vulnerable road users (VRUs), such as pedestrians and cyclists, have not seen the same decline [1]. This leads to the conclusion that today’s passive safety measures can be sufficiently improved if active safety systems are added. Thus, active safety systems need to be tailored to reduce injuries of vulnerable road users.

The key to this goal is an accurate perception of the vehicle’s environment with special interest to the VRU’s behavior. It has been shown that one of the most revealing clues about a person’s intention is the head pose, specifically gaze direction [2]. Also, in urban environments VRUs often appear heavily occluded, thus making it hard to detect the entire body posture while a head can often still be seen in case of common occlusions such as parked cars.

So far, only few studies have addressed the problem of head detection and gaze direction estimation while most focus on the complete body orientation. One can see that this may not be sufficient since information such as a person looking over his shoulder before crossing a street cannot be extracted from a mere body orientation.

This paper addresses the problem of head orientation estimation from high resolution video analysis. We propose a novel framework to estimate a person’s body and head orientation from single frames and integrate the orientation estimation over time from multiple frames. The general outline of our method is depicted in figure 1. In each

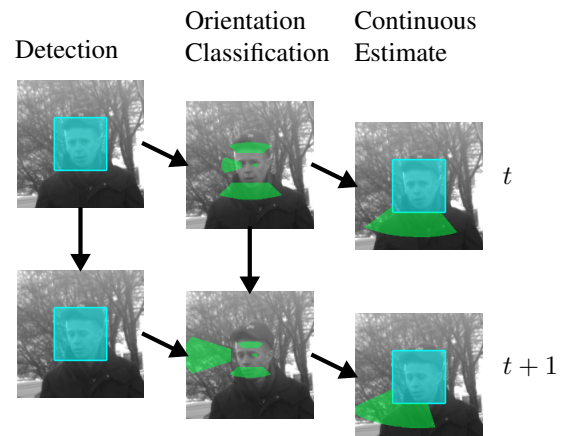


Fig. 1: Hierarchical construction of the processing chain. At each time step we perform detection, discrete orientation classification and then orientation estimation in continuous domain. Detection and orientations are tracked over time.

individual time step a person’s head is detected within the image. Based on this detection four orientation specific classifiers are applied. In the last step, a continuous orientation estimate is generated from the discrete classifications. Over time, detections and orientation estimates are tracked. The remainder of the paper is structured as follows. In section II, we give a short overview over pedestrian detection and orientation estimation in computer vision. In section III we introduce a novel real time capable approach to detect pedestrians heads with their respective orientations. We then evaluate our approach in experiments in section IV and conclude the work in section V.

II. RELATED WORK

In this section gives a brief overview over relevant related work. Since this paper addresses the two problems of detection and orientation estimation, this section is divided into two parts. In section II-A state of the art detection schemes are presented while section II-B focuses on the estimation of body and head orientation.

A. Detection

In recent years pedestrian detection has received great attention throughout the computer vision community, especially in the area of advanced drivers assistance systems. We

*This work supported by the BMWi and the BMW Forschung und Technik GmbH.

¹Eike Rehder and Christoph Stiller are with the Institute of Measurement and Control Systems, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany {eike.rehder, stiller}@kit.edu

²Horst Kloeden is with the BMW Forschung und Technik GmbH, 80992 Munich, Germany horst.kloeden@bmw.de

will give a brief overview over advances pedestrian detection. For a more detailed study, see e.g. [3].

The core concept of most state of the art real time detection is based on the work of Dalal and Triggs [4] who proposed to use a combination of Histogram of Orientated Gradient (HOG) descriptors in combination with a Support Vector Machine (SVM) as a classifier.

To improve classification results, Felzenszwalb *et al.* construct a part-based model from HOG features [5][6]. To detect objects, they apply a root detector at low resolution and then detect parts of the object at double the root resolution while allowing for some deformation of the parts with respect to the root location.

Other than HOG, Viola *et al.* have proposed detection based on boosted cascades using Haar features as classifier input [7]. Also, local receptive fields have been used together with artificial neural networks in order to combine feature extraction and classification [8].

While most approaches detect pedestrians solely in a single modality such as intensity images or optical flow using one single descriptor-classifier combination, Enzweiler *et al.* utilize multiple modalities to compose a mixture of experts system [9]. The proposed system consists of a weighted mixture of different feature and classifier combinations applied to intensity images, optical flow and depth obtained from stereo imaging.

Especially in urban environments, pedestrians often appear part-wise occluded by other objects, e.g. by cars or other pedestrians. To cope with occlusions, Girshick *et al.* extend the detector from [6] to explicitly detect occluded parts within the model.

Especially in high-resolution imaging on which we will not focus in this paper, face recognition is used to detect persons within the image [10]. However, this only applies for faces seen from the front or at maximum in side view. Therefore, this limitation is not sufficient for traffic applications.

B. Orientation Estimation

When it comes to orientation estimation of pedestrians, two different problems have been addressed. The first deals with the estimation of the orientation of a complete persons pose, the second focuses on the orientation of the head. For head orientation estimation, we only focus on medium and low resolution imaging as applicable to vehicle environment perception, unlike for driver monitoring such as [11][12].

For the orientation of the complete pedestrian pose, Gandhi *et al.* use a HOG/SVM stage to classify between eight discrete orientation classes [13]. They also use a Hidden Markov Model (HMM) to stabilize orientation estimations over time.

The part-based detection model proposed by Felzenszwalb [6] can be extended to fix the latent variables within the model to orientation specific models [14]. Thus, responses of the different models within the detection stage automatically yield an assumption of the orientation of a pedestrian in discrete steps.

Enzweiler *et al.* use a similar approach but try to estimate a continuous orientation from the classifier outputs [15]. For this, they again use discrete orientation steps to train their view dependent detectors. The outputs of these classifiers are used as weights for a Gaussian mixture.

Zou *et al.* propose a color based orientation detection scheme [16]. They first track a person's head as an elliptic contour and estimate the orientation by fitting a color model for discrete head orientation classes to the observed features. A continuous orientation angle is obtained from weighting the classes' matching scores.

Chen *et al.* try to extract both head and body orientations as well as motion direction obtained from tracking. They then use a coupling between the three orientations to stabilize head and body orientations as well as a particle filter to track the orientations over time [17][18]. Benfold *et al.* use a similar approach to learn the head orientation in an unsupervised fashion where they train Random Forests from a Conditional Random Field to estimate the orientation based on appearance as well as motion tracking [19]. Recently, Flohr *et al.* have proposed a particle filter to jointly estimate head and body orientation [20].

III. HEAD DETECTION AND ORIENTATION ESTIMATION

Unlike [14][15], we do not estimate the body and head orientation from multiple detector outputs. Instead, we use a classical hierarchical structure which first detects VRUs and then estimates orientations based on these detections (see fig. 1). In the following, the detection stage is explained in section III-A before we introduce our orientation estimation in section III-B. In section III-C integration of the single frame orientation estimation over multiple frames is introduced.

A. Detection

As the focus especially lies on pedestrians in urban environments the detection stage should be able to cope with even heavily occluded pedestrians. Since most occlusions such as parked cars, bushes, etc., only occlude pedestrians to a limited height, we assume that a pedestrian can be detected by just detecting the head.

The detection is based on the HOG/SVM method proposed by Dalal and Triggs [4] as well as the part-based detection scheme proposed by Felzenszwalb *et al.* [6]. In order to meet real time constraints, a cascade based classification is used as depicted in fig 1. The cascade is made up from two stages.

The first stage consists of a simple HOG/SVM classifier using a single head model at low resolution. This detector resembles the one proposed by Dalal and Triggs.

The first stage detector is trained on manually annotated data [21]. The data set consists of roughly 10,000 head images as positives training samples and randomly chosen windows from sequences that do not contain any pedestrians as negative training samples. With our initial detector we perform an exhaustive search for false positives on complete images without any pedestrians. From this, we retrain our final first stage detector. Using an independent test set, the detection threshold is set to a value so that we reduce the

number of false positive detections greatly while keeping as many true positive detections as possible.

The second stage uses a part based model to which is trained only from detections of first stage. While complete pedestrians can vary greatly in pose and appearance, heads only vary little in appearance, e.g. depending on view, hair styles, etc.. However, we found that a part-based detector outperforms a multi-model detector.

The first stage detector is run on negative sequences to extract as many false positives as possible as training data for the second stage. The new training set for the second stage then contains all positive training examples as well as all false positives from the first stage.

We compute the entropy $H(\Omega)$ of the binned image gradients for both orientation and magnitude for the complete training set Ω as well as for the training set after perfect since manually annotated decision with ω for positive and $\bar{\omega}$ for negative samples. The mean information gain $IG(\omega)$ per block is given by blockwise averaging from the pixelwise orientation bin and magnitude features ψ . For the sake of legibility, the pixel coordinates in the following formulas are omitted.

$$H(\Omega) = - \sum_{\psi \in \Omega} P(\psi) * \text{ld}(P(\psi)) \quad (1)$$

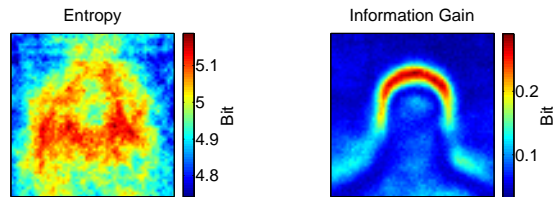
$$IG(\omega) = H(\Omega) - P(\omega)H(\Omega|\omega) - P(\bar{\omega})H(\Omega|\bar{\omega}) \quad (2)$$

Figure 2 shows the pixelwise information gain of the positive samples compared to the negative training samples of the first stage and compared to the false positives after the first stage. In [5] parts were assigned to areas with high positive weights of the root filter. From fig. 2 it can be seen that this might not be the best choice here since true positives and false positives already have a high similarity in regions of high positive weights of the root filter. Note that the entropy of randomly chosen negative samples is almost uniformly distributed over the image while the entropy of false positives shows a strong circular structure similar to a head.

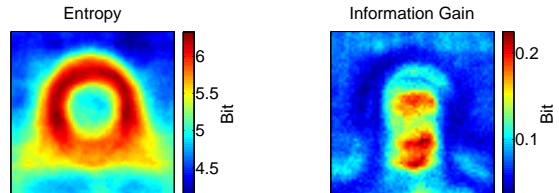
To overcome this problem we focus on parts holding relevant information. We assign three part windows of identical size so that the information density per window is roughly the same and as large as possible. Identical part windows are used since the most computationally demanding task within the classification process is the computation of the descriptor vector.

After assigning sizes and positions of the parts, the corresponding training data is extracted for every part from the training set and train the individual models.

In matching we apply the method proposed by Felzenszwalb *et al.* [5] by computing maps M_i for expected root positions from the part filter responses where $i = 1, \dots, N$ are the respective part indices. To construct the map the part detectors are run on different positions within the bounding box of the first detector stage. From the map of matching scores we compute the map for the expected head center according to (3) where the value for each map entry $M_i(x, y)$



(a) Entropy of randomly chosen negatives and respective information gain.



(b) Entropy of false positives and respective information gain.

Fig. 2: Pixelwise entropy and information gain between positive and different negative samples.

is the matching score of the detector d_i at translated location (x_y, y_d) minus the deformation cost defined by (4).

The deformation cost is modeled as an axis-aligned ellipse displaced from the detection point (x_d, y_d) by the translation between the part position and the root center (x_{c_i}, y_{c_i}) . From the superposition of all parts non-maximum suppression is applied to find the expected center of the head within the bounding box of the first detector stage. We utilize this information to recenter the bounding box and rerun the first stage detector on the recentered bounding box. The final output of the detector is the sum of all part matching scores plus the matching score of the recentered first stage detector d_R .

$$M_i(x, y) = \max_{x_d, y_d} (d_i(x_d, y_d) - w_i(\tilde{x}, \tilde{y})) \quad (3)$$

$$w_i(\tilde{x}, \tilde{y}) = \sqrt{\alpha_i \tilde{x}^2 + \beta_i \tilde{y}^2} \quad (4)$$

$$d(x, y) = d_R + \sum_i M_i(x, y) \quad (5)$$

where $(\tilde{x}, \tilde{y}) = (x_d - x_{c_i} - x, y_d - y_{c_i} - y)$.

After training of first stage and part classifier models the parameters (α_i, β_i) of the uncertainty ellipses in (4) are trained using L2-regularized logistic regression according to

$$e(\alpha, \beta) = \Omega - \tanh(d(\alpha, \beta)), \quad (6)$$

$$J(\alpha, \beta) = e^T e + c(\alpha^2 + \beta^2), \quad (7)$$

$$(\alpha, \beta) = \text{argmin}(J(\alpha, \beta)). \quad (8)$$

In (6), Ω again denotes a vector of the manually annotated labels $\{\omega, \bar{\omega}\}$ of all training data while $d(\alpha, \beta)$ denotes the vector of the detector score computed for all training samples according to (5). The parameter c in (7) weights the regularization in the final cost function and is a parameter to be assigned during training.



Fig. 3: Positive training images with the four views back, left, right and front as well negative training samples.

B. Orientation Estimation

Four discrete orientation classes are used for orientation estimation, i.e. a person facing towards the camera, facing away from it and facing left or right. For every orientation class a classifier is trained to estimate the posterior class probability.

Local Binary Patterns (LBP) serve as features rather than HOG features. The reason for this lies in the contrast normalization within the HOG computation. Due to block-wise normalization in the HOG features even uniform areas appear structured [22]. This makes it extremely difficult to distinguish front and back side of a head from HOG features. To cope with small shifts as well as noise, we construct block-wise histograms of LBP values and concatenate the blocks to one descriptor vector Ψ .

The state vector is defined as the three orientations of head, torso and motion $X = (\Theta_H, \Theta_T, \Theta_M)^T$. The goal is to estimate the joint probability distribution for all possible states. Given the three orientations to be estimated and four orientation classes each we can construct 64 possible state vectors. Due to anatomic limitations all states where head and torso orientations differ by 180° can directly be excluded.

The classifiers for the four different orientation classes are trained using logistic regression. We use logistic regression as the output of each classifier can be interpreted as a posterior probability $P(\Theta|\Psi)$. Other classification methods such as artificial neural networks could also be possible. However, we choose logistic regression over neural networks for runtime reasons.

From manually annotated data the distribution $P(\Theta_H, \Theta_T, \Theta_M)$ is learned as a prior for orientation estimation. Figure 4 shows the distribution of orientation differences between the neighboring state variables. In the training data motion orientation and torso orientation never differ by 180° . With this information the manifold of possible states is reduced to 36.

Due to the use of different detector windows for head and torso orientation classification, the output of the classifiers are statistically independent. Thus the poste-

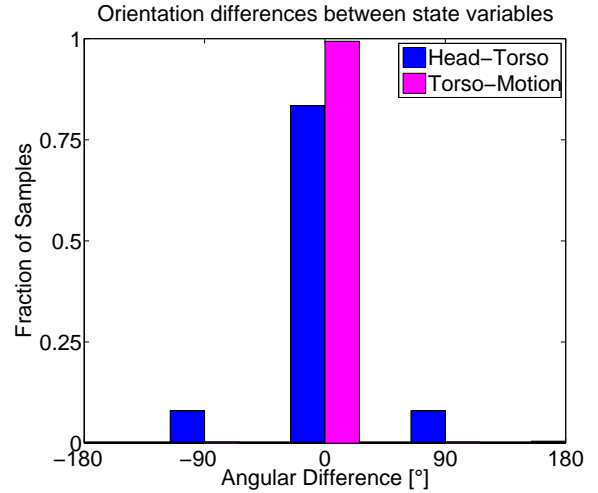


Fig. 4: Difference between individual state variables within same instances.

rior $P(\Theta_H, \Theta_T, \Theta_M|\Psi_H, \Psi_T, \Psi_M)$ can be computed from $P(\Theta_H|\Psi_H)P(\Theta_T|\Psi_T)P(\Theta_M|\Psi_M)$. In case the probabilities cannot be observed, e.g. no torso was detected, a uniform distribution is assumed. In the special case of single frame orientation estimation where there is no motion estimation, the motion orientation in the state vector degenerates to the complete body orientation as addressed in [15].

To obtain the head orientation, we marginalize over all possible states and compute the expectancy of the orientation. With this step, a continuous orientation estimate can be generated from the discrete measurements. From the joint probability distribution we are also able to obtain an assumption over the orientation of the body pose from the measurements of head and torso orientation.

C. Detection and Orientation Tracking

On top of single shot orientation estimation we integrate multiple measurements from video streams into the orientation. In the first step, KLT tracking is applied to the detection bounding boxes [23]. From the estimated motion in image coordinates, motion in world coordinates is computed from camera calibration and an assumption of a standard head size. For this we make use of the fact that human head sizes only vary little compared to for example body sizes [24].

We model the transition of the probability distribution as a Hidden Markov Model of first order with learned state transition probabilities. Given a specific frame rate of our camera we can exclude certain state transitions from further consideration, e.g. a person does not turn his head by 180° inbetween two consecutive frames.

In our model we only estimate the probabilities of discrete states. Therefore a state transition matrix A is employed, containing probabilities learned from manually annotated data as already explained for the priors. In every time step we estimate the prior probability distribution of the state X_t , $P_t^-(X_t|X_{t-1})$ from the previous distribution and the state transition probabilities.



Fig. 5: Images from the test scenes. The blue box depict the detection, the green section depicts the estimated head orientation.

The estimation from the posterior estimates is updated from the classifiers $P(X_t|X_{t-1}, \Psi)$ with $\Psi = (\Psi_H, \Psi_T, \Psi_M)$ with

$$\begin{aligned} \vec{P}_t^-(X_t|X_{t-1}) &= AP_{t-1}(\vec{X}_{t-1}), \\ \vec{P}_t(X_t|X_{t-1}, \Psi) &= H_t(\Psi)P_t^-(X_t|X_{t-1}) \end{aligned} \quad (9)$$

where $H_t(\Psi)$ is the measurement matrix constructed from the classifier outputs.

IV. EXPERIMENTS

A. Data Set

The test data set is a self constructed dataset and consists of eight different scenes shot at different locations [25]. Manually annotated heads in the images serve as a ground truth. In total, the test set features roughly 3,000 frames with more than 2,000 annotated heads in the images.

For evaluation purposes, test persons were equipped with a GPS tracking device as well as an inertia measurement unit (IMU) to measure the head pan angle. Figure 5 shows two samples images from the test set. Note that the test persons are both wearing a baseball cap which was used to mount the IMU. However, the training data mostly features head images of people not wearing caps.

Since the IMU measurements only give the angle relative to the starting point, several discrete orientations were labeled in the image sequences to have a global orientation reference. Also, the labels were used to compensate for the drift of the IMU measurements.

So far, none of the datasets used in related work has been published so that a comparison with other methods can only be done approximately.

B. Results

At first, the detection results are evaluated which are the foundation for any further processing. For the detection experiments a ground plane assumption is used to limit the search space for the detector. Sliding window detection is

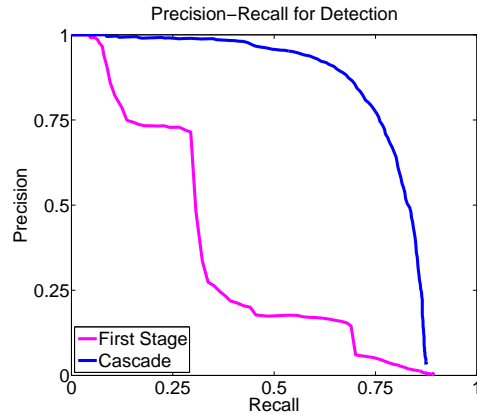


Fig. 6: Precision vs. recall for detection after the first stage (magenta), after the second stage (blue)

performed on an image pyramid of the region of interest. The search stride for the first stage is a quarter of the window size while in the second stage it is an eighth of the window size.

Figure 6 shows the precision-recall curve of the detection after both stages of the cascade as well as after tracking. It can be seen that the use of the part-based cascade significantly boosts the detector performance from an average precision of 33% for the single stage to 78% after the second cascade stage. Still, the detector remains capable of real time application on a standard 2.6GHz quad-core laptop. Note that after the first stage we have already assigned a decision threshold for the previous stage, i.e. the magenta curve corresponds to the precision-recall curve of the HOG/SVM classification from [4] whereas the blue curve corresponds to the precision-recall of the second stage at one fixed decision value for the first stage. By setting a decision threshold for the first stage we also assign an upper bound for the recall of the cascade to the recall of the first stage. For evaluation of the orientation estimation only detections are used that can be associated to an annotated head in the image. Figure 7 shows the orientation estimate (blue) and the ground truth data obtained from the IMU (green) for one sequence. In this sequence, the test person rotated on the spot first clockwise and then switched the direction to counter clockwise rotation. It can be seen from the curve that the orientation estimate is close to the ground truth measurement and also is able to cope with the direction change instantaneously. We evaluated the orientation estimation over all detections for which an IMU orientation measurement was available. The error distribution can be seen in figure 8. The histogram shows a strong peak at zero degrees orientation error while errors at around 180° are almost non present. In total we reach an average absolute error of 19°. To our best knowledge, this is the lowest error to be achieved in head orientation estimation. However, no exact comparison is possible as no unified dataset exists.

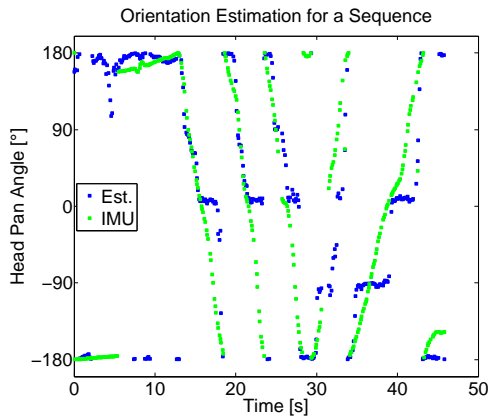


Fig. 7: Orientation Estimate (blue) and ground truth (green) for a complete sequence.

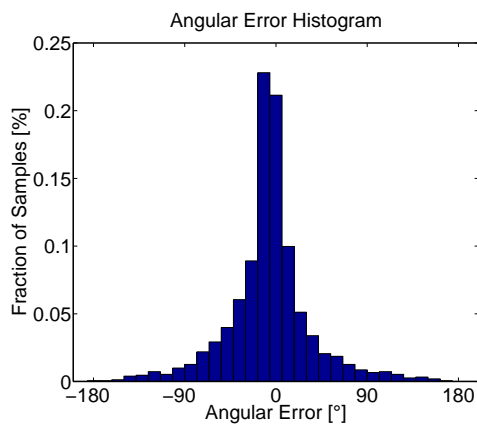


Fig. 8: Orientation estimate errors. Mean absolute error: 19° .

V. CONCLUSIONS

In this paper we have proposed a novel framework to detect and estimate the orientation of heads of pedestrians in traffic scenarios.

For this we have utilized a part based HOG/SVM cascade for detection. Our findings in the detection stage suggest the need for features orthogonal to HOG features to surpass current detection algorithms. Thus future work should focus on the development of real time applicable feature enhancement.

For the orientation estimation we have used a L_2 -regularized logistic regression model together with Bayesian filtering of discrete orientation classes. We have shown that our approach is capable of an accurate continuous orientation estimation.

Further improvement could be reached when eight rather than four orientation classes are used for the orientation classification. However, the results with four orientation classes already show a significant improvement over other proposed methods using eight classes.

REFERENCES

- [1] Statistisches Bundesamt, “Verkehrsunfälle 2013,” vol. 8, no. 7, April 2014.
- [2] S. Schmidt, B. Färber, and A. Pérez Grassi, “Geht er oder geht er nicht?—Ein FAS zur Vorhersage von Fußgängerabsichten,” in *Workshop Fahrerassistenzsysteme, Freundeskreis Mess- und Regelungstechnik Karlsruhe eV, Walting*, 2008, pp. 2–4.
- [3] M. Enzweiler and D. Gavrila, “Monocular pedestrian detection: Survey and experiments,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 12, pp. 2179–2195, Dec. 2009.
- [4] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” vol. 1, Jun. 2005, pp. 886–893 vol. 1.
- [5] P. Felzenszwalb, D. McAllester, and D. Ramanan, “A discriminatively trained, multiscale, deformable part model,” in *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, Jun. 2008, pp. 1–8.
- [6] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627–1645, 2010.
- [7] P. Viola, M. Jones, and D. Snow, “Detecting pedestrians using patterns of motion and appearance,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, Oct 2003, pp. 734–741 vol.2.
- [8] C. Wöhler and J. Anlauf, “An adaptable time-delay neural-network algorithm for image sequence analysis,” *Neural Networks, IEEE Transactions on*, vol. 10, no. 6, pp. 1531–1536, Nov 1999.
- [9] M. Enzweiler and D. Gavrila, “A multilevel mixture-of-experts framework for pedestrian classification,” *Image Processing, IEEE Transactions on*, vol. 20, no. 10, pp. 2967–2979, Oct. 2011.
- [10] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” vol. 1, 2001, pp. 1–511 – 1–518 vol.1.
- [11] Y. Zhao and H. Yan, “Head orientation estimation using neural network,” in *2011 International Conference on Computer Science and Network Technology (ICCSNT)*, vol. 3, Dec. 2011, pp. 2075–2078.
- [12] S. J. Lee, J. Jo, H. G. Jung, K. R. Park, and J. Kim, “Real-time gaze estimator based on driver’s head orientation for forward collision warning system,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 254–267, Mar. 2011.
- [13] T. Gandhi and M. Trivedi, “Image based estimation of pedestrian orientation for improving path prediction,” in *2008 IEEE Intelligent Vehicles Symposium*, Jun. 2008, pp. 506–511.
- [14] A. Geiger, C. Wojek, and R. Urtasun, “Joint 3d estimation of objects and scene layout,” in *NIPS*, 2011.
- [15] M. Enzweiler and D. Gavrila, “Integrated pedestrian classification and orientation estimation,” Jun. 2010, pp. 982–989.
- [16] W. Zou, L. Fang, Y. Li, and K. Yuan, “Head orientation estimation for covert-tracking robot,” in *2009 IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, Dec. 2009, pp. 322–327.
- [17] C. Chen, A. Heili, and J.-M. Odobez, “A joint estimation of head and body orientation cues in surveillance video,” in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Nov. 2011, pp. 860–867.
- [18] C. Chen and J. Odobez, “We are not contortionists: Coupled adaptive learning for head and body orientation estimation in surveillance video,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2012, pp. 1544–1551.
- [19] B. Benfold and I. Reid, “Unsupervised learning of a scene-specific coarse gaze estimator,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*, Nov 2011, pp. 2344–2351.
- [20] F. Flohr, M. Dumitru-Guzu, J. Kooij, and D. Gavrila, “Joint probabilistic pedestrian head and body orientation estimation,” in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, June 2014, pp. 617–622.
- [21] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, “LIBLINEAR: A library for large linear classification,” *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.
- [22] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba, “Hoggles: Visualizing object detection features,” in *Computer Vision (ICCV), 2013 IEEE International Conference on*, Dec 2013, pp. 1–8.
- [23] J. Shi and C. Tomasi, “Good features to track,” in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR ’94., 1994 IEEE Computer Society Conference on*, Jun 1994, pp. 593–600.
- [24] NASA, “NASA-STD-3001 man-systems integration standards,” July 1995.
- [25] H. Klöden, N. Brouwer, S. Ries, and R. Raschofer, “Potenzial der Kopfposenerkennung zur Absichtsvorhersage von Fußgängern im urbanen Verkehr,” in *FAS Workshop Fahrerassistenzsysteme, Walting, Germany*, 2014.