Early Pedestrian Movement Detection Using Smart Devices Based on Human Activity Recognition

Diego Botache, Liu Dandan, Maarten Bieshaar, Bernhard Sick[1]

Abstract: In the future, vulnerable road users (VRUs) such as cyclists and pedestrians will be equipped with smart devices capable of communicating with intelligent vehicles and infrastructure. This allows for cooperation between all traffic participants, such as cooperative intention detection and future trajectory prediction for advanced VRU protection. Smart devices can be used to detect the pedestrians' intentions to warn approaching vehicles. In this article, we propose a method based on human activity recognition for early pedestrian movement transition detection using smart devices. These movement detections serve as valuable information for pedestrian path prediction and intention detection. We represent the pedestrians' behavior using four states, i.e., waiting, starting, moving, and stopping. The movement transition detection is modeled as a classification problem and tackled by means of machine learning classifiers. The labels for training the classifier are obtained by evaluation of recorded high-precision head trajectories. We compare two different classification paradigms: A simple support-vector machine with linear kernel and a more complex XGBoost classifier. Our empirical studies with real-world data originating from experiments which 11 test subjects involving 79 different scenes show that we are able to detect movement transitions robust and early, reaching an F_1-score of 85%.

Keywords:

Vulnerable Road Users; VRU safety; VRU Intention Detection; Cooperative Intention Detection; Artificial Intelligence; Machine Learning; Pedestrian Movement Detection; Human Activity Recognition

1 Introduction

1.1 Motivation

In our work, we envision the following mixed traffic scenario in which intelligent, automated vehicles, trucks, sensor-equipped infrastructure, and vulnerable road users (VRUs), such as pedestrians and cyclists, equipped with smart devices (e.g., smartphones and other wearables) are interconnected by means of an ad-hoc network. The collective intelligence of all road users is used to determine and maintain a cooperative model of the current (traffic) environment [Bi17a]. To avoid accidents involving vehicles and VRUs, it is not only important to detect VRUs but also to anticipate their intentions. Modern vehicles are equipped with forward looking active safety systems (e.g., radar), nevertheless, dangerous situations involving VRUs may occur as a result of sensor malfunctions or occlusions.

[1] University of Kassel, Intelligent Embedded Systems, Wilhelmshöher Allee 73, 34121 Kassel, Germany
{diego.botache, ldandan, mbieshaar, bsick}@uni-kassel.de

doi:10.18420/inf2019_ws26

Figure 1 shows a typical critical occlusion situation involving a pedestrian intending to cross the street while a vehicle hidden behind the bus is approaching. Smart devices worn by the pedestrian could anticipate the pedestrian's intention to cross the street and transmit it to the approaching car. Then, a driver warning or automated emergency breaking maneuver can be issued avoiding a potentially fatal accident. The challenge is to detect a movement transitions fast and yet reliable. For illustration consider the following example of an urban scenario: An automated vehicle is approaching with 50 km/h and has a braking deceleration of 8 m/s^2. If the breaking maneuver is initiated 15 m ahead of the crossing pedestrian, the car will come to a standstill 3 m before the pedestrian. After the pedestrian has entered the driving corridor, the control system of the vehicle has 0.58 s to initiate a braking maneuver and to avoid an accident. Nevertheless, the detector must also be robust, i.e., avoid false positive detections potentially leading to unnecessary emergency braking maneuvers.

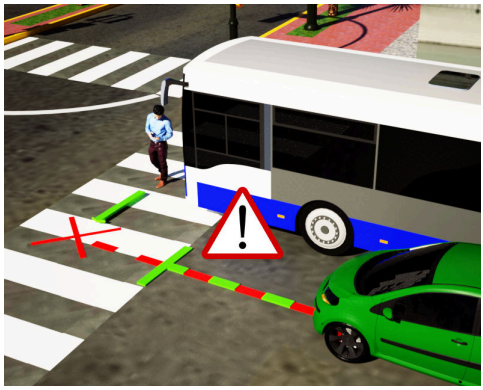


Fig. 1: An example of a dangerous situation in an urban environment. A pedestrian is intending to cross the street occluded by the bus such that it cannot be seen by the approaching car. The smart device worn by the pedestrian anticipates the pedestrian's intention and transmits a warning to the approaching car such that a potential collision is avoided.

Modern smart devices are equipped with integrated global navigation satellite system (GNSS) receivers and inertial sensors (i.e., accelerometer and gyroscope). In contrast to GNSS-based sensing, inertial sensors are not affected by GNSS outage (often encountered in urban environments) and, moreover, are available at a high sample rate, enabling the detection of fast movement transitions, i.e., within half a second. Our approach to early detection of pedestrian movement's is based on human activity recognition (HAR) [BBS14] and machine learning techniques using inertial sensors, only. The focus of this article is the early and robust movement detection of pedestrians. The results can be used to support trajectory forecasting, e.g., as presented in [Bi17b].

1.2 Main Contribution

Our main contribution is an approach based on HAR and machine learning using inertial sensors to detect pedestrian movements. The approach is based on an adaption of the cyclist's starting movement detection approach using smart device as presented in [Bi18a]. The movement detection is modeled as a classification task. In this article, we consider the following aspects: first, modeling of the pedestrian's behavior using four states and

the detection involving multi-class classification based on HAR. Second, a comprehensive evaluation of the approach for pedestrian movement detection, showing promising classification performance. These results can then be used to improve pedestrian's trajectory forecast [Go16] and ultimately increase safety. In this article, we considered an off-line evaluation of our approach with real data. Aspects concerning algorithmic runtime and processing time are neglected.

The remainder of this article is structured as follows: In Section 2 the related work is reviewed. In Section 3, the general pedestrian behavior modelling, the method for ground truth label generation, and the approach to detect pedestrians' movements are detailed. In Section 4, the evaluation methodology as well as the data acquisition is presented. In Section 5, the experimental results are reviewed before finally in Section 6, a conclusion is drawn and possible directions of future work are sketched.

2 Related Work

As it was shown in [Bi17b], early knowledge of the movement of VRUs can support the trajectory forecast. Most of the research in intention detection relies on vision-based solutions, e.g., [KG14, Ko14] the authors showed a promising approach for vehicle-based pedestrian detection and short-term forecasting using cameras. Vision-based approaches require line of sight and fail in the presence of sensor-outage and occlusions. Yet, many dangerous situations occur as a result of occlusion. Cooperative intelligent transportation systems (C-ITS) aim to overcome this shortcoming by means of cooperation between the different vehicles including smart devices worn by the VRUs [SvSM17]. Using Car2Pedestrian (C2P) communication, vehicles and smart devices carried by the VRUs can cooperate [FD09], i.e., the smart device can share the current VRU's position, velocity, and heading to avoid potential collisions. In [En13], a system relying on C2P communication for tracking pedestrians was proposed. It combines GNSS data with inertial sensors allowing to transmit position and movement type to an approaching car. A smartphone-based collision avoidance system is proposed in [BMD17], where additional context information obtained from a pedestrian's smart device is used to improve the collision detection accuracy. A collision avoidance system based on 5G/LTE given in [JMD18] identifies VRUs in potentially dangerous situations based on information about the location and movement direction. An approach to cooperative perception and intention detection of cyclists including smart devices is presented in [Bi17b] and [Bi18b]. Here, we will present the basis for the extension of these cooperative approaches to also cover pedestrians.

3 Methodology

We aim at detecting pedestrian's movement transitions using smart devices carried by the pedestrian. The pedestrian's movements are modeled using four different states, i.e.,

'waiting', 'starting', 'moving', and 'stopping'. Based on these states, we define four distinct classes for which we obtain labels by inspection of the pedestrian's head trajectory [Go16]. Subsequently, the movement detection is modeled as a classification task. Whereas the movement detector, which aims to detect the current movement state, is realized by means of a HAR pipeline [BBS14] using the inertial sensors of the smart device, i.e., accelerometer and gyroscope as input.

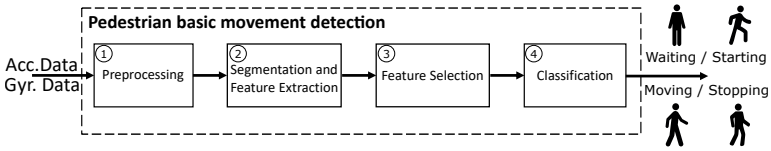


Fig. 2: Process for basic movement detection of pedestrians based on smart devices

The HAR pipeline underlying our approach to detect the pedestrian's movement is depicted in Fig. 2. It consists of four steps: first, preprocessing of the acceleration and gyroscope data, second, segmentation and feature extraction, third, feature selection and dimensionality reduction, and finally, classification by means of machine learning classifiers, i.e, support-vector-machine with linear kernel (linear SVM) and extreme gradient boosting classifier (XGBoost) [CG16].

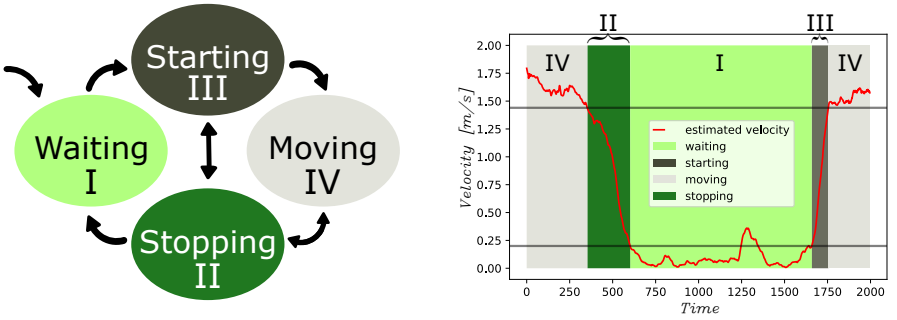
In the remainder of this section, we first introduce the pedestrian's behavior model and the automated labeling procedure. Subsequently, we present the four steps of our approach.

3.1 Pedestrian Behavior Model and Class Labeling

We model the movement of pedestrians using a state machine consisting of four different states [Go16]. A schematic of this state machine is depicted in Fig. 3 (a). Transitions in the state machine correspond to changes in the movement dynamics of the pedestrian's behavior. As it was shown [Go16], the integration of this four-state behavior model can help to significantly improve the intention detection performance, i.e, the pedestrian trajectory forecasting quality. The movement states of the pedestrian are defined via the velocity derived from the pedestrian's trajectory. In Fig. 3 (b), there is an example of velocity which is ordered according to the four different movement states. The definition of these movement states, i.e., the labeling of the classes, follows the approach presented in [Go16]. It uses thresholds defined on the absolute velocity derived from the pedestrian's head trajectory. Moreover, it is based on the observation that pedestrians typically possess a steady-state velocity. In order to get high quality labels, i.e., avoid cluttered segmentation labels, the trajectory is smoothed using a mean filter. An offline processing is necessary to compensate for the delay introduced by the mean filtering of the input signal. Waiting is defined as the time segments for which the pedestrian's velocity is below a predefined waiting threshold, which is set to 0.2 m/s [Go16]. Waiting is followed by the starting movement, i.e., velocity exceeds the waiting threshold. The end of the starting movement and the transition to moving are defined by means of a moving threshold. The user dependent moving threshold,

which is motivated by the steady-state velocity of pedestrians, is set to 80 % of the maximal velocity of the pedestrian within the considered experiment. Stopping follows moving and is defined as the velocity falling below the moving threshold. Yet, a movement is only labeled stopping if it is followed by a waiting (i.e., velocity falls below the waiting threshold). If this is not the case, the potential stopping segment is labeled as moving. Hence, detecting a stopping movement can also be considered as a forecasting task.

The definition of the movement states via the velocity cannot be directly used for detection using smart devices, as it involves offline processing methods. Therefore, in the following we present our approach based on HAR which aims to detect the four states by means of classification.



(a) State machine of the pedestrian model

(b) Labeling

Fig. 3: Labeling of the pedestrian states based on head trajectories. Waiting: The pedestrian is standing in a place(moving of the upper body or head is possible). Starting: The pedestrian starts to move from a waiting status and reaches a steady-state velocity. Moving: The pedestrian is moving or walking, continually. Stopping: The pedestrian starts to decelerate and finally stops.

3.2 Preprocessing, Feature Extraction, and Feature Selection

Our approach uses the smart device's inertial sensors, i.e., data originating from gyroscope and gravity compensated accelerometer measurements sampled with 50 Hz. The three components (x, y, and z) of both sensors are transformed using the orientation estimation supplied by the smart device's operating system. The resulting local tangential coordinate frame is leveled with the local ground earth plate, i.e., the z-axis is pointing towards the sky. We avoid the tedious estimation of the transformation between the device and the pedestrian orientation by only considering the magnitude of the linear accelerometer and gyroscope data in the local tangential horizontal x-y plane. This representation is invariant concerning the smart device orientation. Additionally, we also consider the projection of the accelerometer and gyroscope measurements on the z-axis of the local tangential frame. In total, we have four distinct input signals. GNSS data has not been used because of two reasons: first, lack of signal coverage and multipath effects in urban areas, second, low

sampling frequencies of modern smart devices (1 Hz) which do not allow to detect fast dynamic changes in the pedestrian's movement.

Based on the inputs, we perform a sliding window segmentation for which we then compute standard HAR features [BBS14], e.g., energy, minimum and maximum. To handle dynamics on different time scales, we consider four different window sizes, i.e., 0.1 s, 0.5 s, 1.0 s, and 2 s.

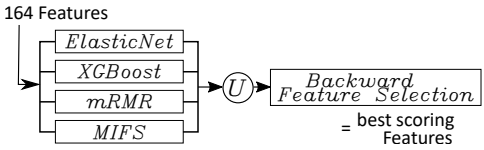


Fig. 4: Two-stage feature selection procedure.

Moreover, we also consider features based on orthogonal polynomial approximations (i.e., best estimators of mean, slope, and curvature) up to the third degree for window sizes of 0.5 s, 1.0 s, and 2.0 s. In addition to that, the discrete Fourier Transform (DFT) of two separated window sizes of 5,12 s and 0,64 s up to 5th degree (as human movement is well described by lower frequencies [Pa12]) are considered. To make the DFT coefficients independent of the energy, a normalization is required. In total, 164 features are computed.

We applied a two-stage feature selection for obtaining robust and human understandable features [Bi18a]. A schematic of this is depicted in Fig. 4. In the first stage, four feature extraction filters, i.e., mutual information (MIFS), minimum redundancy maximum relevance (mRMR), filters based on ElasticNet, and XGBoost, are applied. The ten best scoring features are unioned. In total, maximal 40 features are selected. In the second stage, a feature selection refinement using a backward feature selection is applied. Here, features are selected to optimize the F_1 score in conjunction with the respective classifier, i.e., SVM with linear kernel or XGBoost classifier. A comprehensive and detailed description of the preprocessing, feature extraction and selection can be found in [Bi18a].

3.3 Classification

The detection of pedestrian movement is realized by means of a frame-based XGBoost [CG16] and SVM with linear kernel classifier. The XGBoost algorithm combines the optimization of an arbitrary differentiable loss function for consideration of the training loss with an additional regularization to reduce model variance. Additional regularization techniques based on shrinkage of the learning rate and random feature sub-sampling further enhance the generalization ability. The linear SVM is selected because of its simplicity and good generalization, i.e., classification performance in many applications. The Hinge loss is applied to optimize the linear SVM classifier and due to a large number of training samples, the linear SVM is trained in the primal space. Frame-based classification is implemented at the discrete points at a 50 Hz frequency.

Both classifier are trained based on labeled data with the four pedestrian movement classes. The classes are highly imbalanced, i.e., starting and stopping classes are underrepresented. A

resampling strategy based on [BPM04] known as SMOTETomek compensates imbalanced classes combining under- and oversampling. The effective class ratio is an important factor influencing the result of the movement classification. In our approach, we use the results of the classification to calculate posterior class probabilities. These probabilities represents confidence estimates about the classification of each movement state. Based on these probability estimates, the classification can be derived, i.e., class with highest probability. Well-calibrated probability estimates represent the detectors confidence and are important for the fusion of different detection results, e.g., combination of detections from vehicles and smart devices for cooperative perception in the envisioned future traffic scenario. In addition, the probabilities are also used for trajectory prediction [Bi17b]. Therefore, a probability calibration using a Platt scaling is performed.

4 Data Acquisition and Evaluation Methodology

For evaluation of our approach, we consider a dataset consisting of pedestrian movements. It contains 11 female and male test subjects. The test subjects were instructed to move between predestined points at an urban intersection with public, uninstructed traffic. They were obliged to obey the traffic rules. While the pedestrians moved across the chosen intersection located at the city of Aschaffenburg, Germany, their movements, i.e., trajectories, were recorded by a high-resolution wide-angle stereo camera system [Go12]. To obtain the head trajectory for labeling a pedestrian's movement, we manually annotated the pedestrian's head detections on both cameras. The 3D position of the head is obtained by triangulation of the annotated positions. Based on this head trajectory, we applied the automated labeling procedure as described in Section 3.1. In total, the data comprises a length of 38 minutes including 79 scenes.

The smart devices used during the experiments are Samsung Galaxy S6 smartphones. The test subjects were equipped with a smartphone in their left front trouser pocket. The smartphone was placed in an upright position with the screen facing outwards. Note, that this setup aims to increase the reproducibility of our experimental setting and does not limit the general applicability of our approach with respect to potential other wearing locations [Bi18a].

The evaluation of our movement detection approach is performed offline using a ten-fold cross-validation over the test subjects. For the purpose of comparing the performance of the two classifiers, a number of scores including F_1-score, accuracy, precision, and recall are evaluated. Moreover, we considered the confusion matrix for evaluation.

5 Experimental Results

This section presents the results for the SVM and XGBoost classifiers. We evaluate the detection performance for 250 random parameter combinations by means of a ten-fold

cross-validation over the test subjects. For the XGBoost classifier, we considered the following parameter configuration: Number of trees (50, 100, 200, 300, 500, and 700), the maximum tree depth (between 3 and 10), and the learning rate (between 0.01 and 0.2). For the linear SVM we only considered the penalty term (between 2^{-8} and 2^{8}). A random subsampling was performed in order to speed up the training and evaluation process. We considered three classifiers. One linear SVM classifier and XBoost classifier (XGBoost classifier F_1) based on the highest validation F_1-score and one XBoost classifier (XGBoost classifier conf.) based on inspection of the confusion matrices.

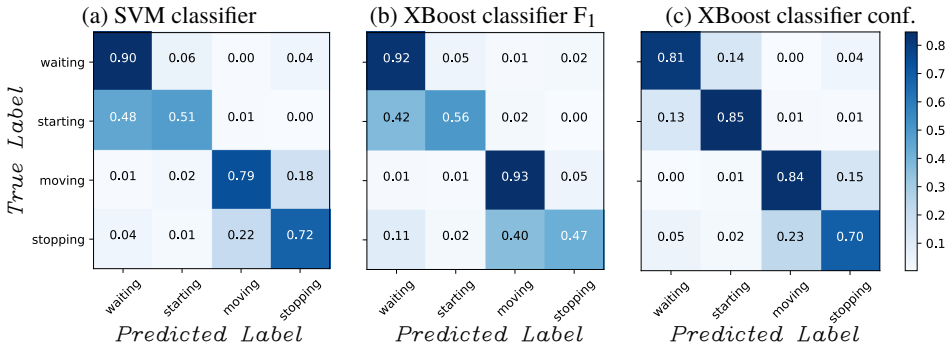


Fig. 5: Confusion matrix for classification precision of each class shown at the diagonal. Both classifiers with the highest F1-Score (a) and (b) have more than 90 % precision for waiting. (c) The XGBoost classifier conf. reaches for the class waiting a precision of 81 % but has the best precision for the class starting with a value of 85 %, where the selected SVM classifier (a) only shows a 51 % precision and the selected classifier of XGBoost at (b) gives a low precision of 56 %. The best result for the class moving is given at (b) with a precision of 93 %. Furthermore, we observe the best result for the classification accuracy for the class stopping with the SVM classifier, which reaches a value of 72 %.

The first two classifiers (a) and (b) shown in Fig. 5 were selected according to the maximum F_1-score. (a) The SVM classifier with the sweeping parameter 2.0 reaches an F_1-score values of 85.6 %. (b) The selected XGBoost classifier reaches an F_1-score of 88,9 %. The parameters are 200 tree, a maximal depth of 7 and a learning rate set to 0.041. (c) The selected XBoost classifier based on the confusion matrix reaches an F_1-Score of 84,98 %.

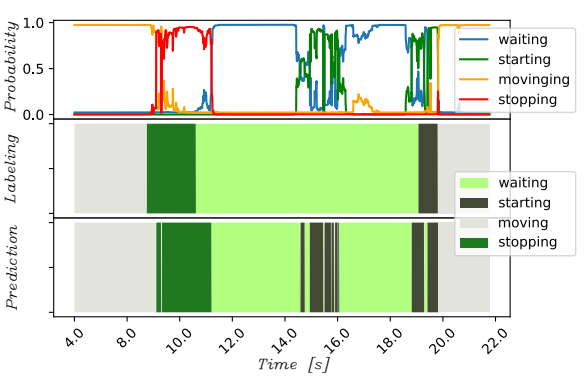


Fig. 6: Selected Scene

In the following, we consider a sample scene of the latter XBoost classifier. It is depicted in Fig. 6. Blue, green, orange, and red lines represent the probability assigned to waiting,

starting, moving, and stopping separately in the first plot. The labeling plot represents the ground truth data as given in section 3.1 and the prediction plot gives the state of the highest class value at certain time.

6 Conclusions and Future Work

In this article, we have presented an automated procedure for data labeling based on [Go16] and a pedestrian behavior model, which has been used for training and evaluation of linear SVM and XGboost classifiers. The approach is based on HAR pipeline with a two-stage feature selection procedure. In experiment including 79 Scenes with 11 pedestrians, we compared different classifiers, including a linear SVM and different XGBoost classifiers. The selected XGBoost classifier reaches a F_1-Score of 88,9%. Yet, the detailed evaluation of the confusion matrix hints that only the consideration of the F_1-score is not sufficient for selecting fast and robust movement detectors. Nevertheless in the following, we are going to integrate the presented early movement detection approach for improved pedestrian's trajectory forecast [Bi17b].

The approach presented in this article, is a step towards our envisioned future traffic scenario involving cooperative intention detection [Bi17a]. Fusion methods comprising information originating from smart devices and vehicles to identify potential dangerous situations [JMD18] shall be investigated in future work. Moreover, we will considered the integration of a different user-centric coordinate system for feature extraction [Ja17] for improved movement detection.

Acknowledgment

This work results from the project DeCoInt^2, supported by the German Research Foundation (DFG) within the priority program SPP 1835: "Kooperativ interagierende Automobile", grant number SI 674/11-1.

Bibliography

[BBS14] Bulling, A.; Blanke, U.; Schiele, B.: A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors. ACM Comput. Surv., 46(3):1–33, 2014.

[Bi17a] Bieshaar, M.; Reitberger, G.; Zernetsch, S.; Sick, B.; Fuchs, E.; Doll, K.: Detecting Intentions of Vulnerable Road Users Based on Collective Intelligence. In: AAET – Automatisiertes und vernetztes Fahren. Braunschweig, Germany, pp. 67–87, 2017.

[Bi17b] Bieshaar, M.; Zernetsch, S.; Depping, M.; Sick, B.; Doll, K.: Cooperative Starting Intention Detection of Cyclists Based on Smart Devices and Infrastructure. In: ITSC. Yokohama, Japan, 2017.

[Bi18a] Bieshaar, M.; Depping, M.; Schneegans, J.; Sick, B.: Starting Movement Detection of Cyclists using Smart Devices. In: International Conference on Data Science and Advanced Analytics (DSAA). Turin, Italy, pp. 1–8, 2018.

[Bi18b] Bieshaar, M.; Zernetsch, S.; Hubert, A.; Sick, B.; Doll, K.: Cooperative Starting Movement Detection of Cyclists Using Convolutional Neural Networks and a Boosted Stacking Ensemble. CoRR, arXiv:1803.03487, 2018.

[BMD17] Bachmann, M.; Morold, M.; David, K.: Improving smartphone based collision avoidance by using pedestrian context information. In: PerCom Workshops. Kona, HI, pp. 2–5, March 2017.

[BPM04] Batista, G. E. A. P. A.; Prati, R. C..; Monard, M. C.: A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. SIGKDD Explor. Newsl., 6(1):20–29, June 2004.

[CG16] Chen, T.; Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In: KDD16. San Francisco, CA, pp. 785–794, 2016.

[En13] Engel, S.; Kratzsch, C.; David, K.; und M. Holzknecht, D. Warkow: Car2Pedestrian Positioning: Methods for Improving GPS Positioning in Radio-Based VRU Protection Systems. In: 6. Tagung Fahrerassistenzsysteme. Munich, Germany, 2013.

[FD09] Flach, A.; David, K.: A Physical Analysis of an Accident Scenario between Cars and Pedestrians. In: VTC Fall. Anchorage, AK, pp. 1–5, 2009.

[Go12] Goldhammer, M.; Strigel, E.; Meissner, D.; Brunsmann, U.; Doll, K.; Dietmayer, K.: Cooperative Multi Sensor Network for Traffic Safety Applications at Intersections. In: ITSC. Anchorage, AK, pp. 1178–1183, 2012.

[Go16] Goldhammer, M.: Selbstlernende Algorithmen zur videobasierten Absichtserkennung von Fußgängern. Intelligent Embedded Systems. Kassel University Press, 2016. (Dissertation, Universität Kassel, Fachbereich Elektrotechnik/Informatik).

[Ja17] Jahn, A.; Bachmann, M.; Wenzel, P.; David, K.: Focus on the User: A User Relative Coordinate System for Activity Detection. In: Modeling and Using Context. Paris, France, pp. 582–595, 2017.

[JMD18] Jahn, A.; Morold, M.; David, K.: 5G Based Collision Avoidance - Benefit from Unobtrusive Activities. In: 2018 European Conference on Networks and Communications (EuCNC). pp. 1–356, June 2018.

[KG14] Keller, C. G.; Gavrila, D. M.: Will the Pedestrian Cross? A Study on Pedestrian Path Prediction. TITS 13, 15(2):494–506, 2014.

[Ko14] Kooij, J.; Schneider, N.; Flohr, F.; Gavrila, D.: Context-based pedestrian path prediction. In (Fleet, David; Pajdla, Tomas; Schiele, Bernt; Tuytelaars, Tinne, eds): ECCV 2014. Zürich, Switzerland, pp. 618–633, 2014.

[Pa12] Park, J.; Patel, A.; Curtis, D.; Teller, S.; Ledlie, J.: Online Pose Classification and Walking Speed Estimation Using Handheld Devices. In: UbiComp. New York, NY, pp. 113–122, 2012.

[SvSM17] Scholliers, J.; van Sambeek, M.; Moerman, K.: Integration of vulnerable road users in cooperative ITS systems. ETRR, 9(2):15, 2017.