Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Prediction of driver behavior on a limited sensory setting

Michael Garcia Ortiz, Jens Schmüdderich, Franz Kummert

2012

Preprint:

This is an accepted article published in Intelligent Transportation Systems Conference (ITSC). The final authenticated version is available online at: https://doi.org/[DOI not available]

2012 15th International IEEE Conference on Intelligent Transportation Systems
Anchorage, Alaska, USA, September 16-19, 2012

Prediction of driver behavior on a limited sensory setting

Michaël Garcia Ortiz, Franz Kummert and Jens Schmüdderich

Abstract— **We present a system able to predict the future left lane change behavior of the ego-vehicle on highway, depending on detected information about the preceding vehicles. Our system learns the mapping between the current perceived scene (information about the ego-vehicle and the preceding vehicles) and the future driving behavior of the ego-vehicle.**

The main difference to related contributions comes from the fact that we want to perform behavior prediction with a limited sensory setting. On the one hand, we do not perform driver monitoring, and on the other hand, we only use frontal sensors. However, we show that it is possible to predict accurately the left lane change behavior of the driver.

The system is tested on a real world scenario, using 50 minutes of streams recorded on German highways, and the prediction occurs at a horizon of 2s into the future. The quality of the prediction is measured using frame-based as well as event-based evaluation with a cross-validation approach. We compare a system with only front sensing (vision and radar detection) with a system including a simulated rear sensing.

I. Introduction

Recent contributions in the domain of Intelligent Transportation Systems propose advanced sensory setting involving, e.g., 360 degrees laser detection. The increase in capability of online computation modules allows the use of elaborated sensory processing algorithms online (as, e.g., SLAM on laser points [1], on-the-fly object detection [2], free area estimation [3]). These combinations of hardware and techniques lead to impressive results, and can be used to provide a scene representation for Driver Assistant Systems.

An important step in the development of Driver Assistant Systems is driver behavior prediction: it is necessary to anticipate what the driver usually does to predict his future behavior. By comparing the prediction with the observation, a prediction system can detect if an unusual behavior occurs. The risk related to a scene can be estimated or predicted in order to take action or warn the driver soon enough (e.g. [4], [5], [6], [7] or [8]). A system which is only reactive will be too late to prevent accidents.

The main issue with elaborated sensory systems is the fact that these technologies come at a high cost: they are expensive, and the computational power needed is enormous. Not all cars can be equipped with such systems, so the technology will only be accessible for a limited amount of people. Obviously, the overall benefit of this technology will be drastically reduced: a driving assistant system which is used by few people will prevent few accidents. On the contrary, a cheaper driving assistant system, even less precise, but accessible to more drivers, might prevent more accidents. For these reasons we want to explore the possibilities of behavior prediction on a limited sensory setting. How good can a behavior prediction system get when we deliberately reduce the quantity of sensors?

In this contribution, we present a lane-change prediction system on highway based only on front sensing. We chose to limit the sensory setting in order to evaluate possible applications for mobile devices. One of these lane changes is presented in Fig. 1 The presented system should solve a classification task answering the following question: Is the driver going to perform a lane change in the next 2 seconds? We avoid using any form of driver monitoring (like blinkers, head orientation, or gaze) for reasons explained in Sec. II.

We present several indicators representing behaviorally relevant knowledge. We show that front sensing is enough to give satisfactory result, and that the front scene holds most of the information necessary to predict the lane change behavior. Of course, we also show that, as expected, additional information about the rear scene helps the system to achieve better prediction. We test our system using event-based as well as frame-based evaluation, and we use N-Fold cross validation in order to ensure a reliable and fair evaluation.

II. Related Work and novelty of the contribution

Numerous contributions these past few years presented systems to infer the driver future behavior ([9], [10], [11], [12], [13] or [14]). We present three contributions that we consider related to the approach we propose in this study.

In [15] the authors present a study of indicators for lane-change intent extracted by observing the surroundings as well as by monitoring the driver. Even if they do not perform prediction, they conclude their study by affirming that the driver head monitoring and the left turn signal are strong indicators for lane change intent.

In [16], the authors equipped a car with limited sensory device: one front camera and one rear camera. They also use driver monitoring in order to extract the position, orientation and viewpoint of the driver. They can predict different types of driving behaviors on highway on average 1 second before they occur, and in particular the lane change on average 0.1 seconds before it occurs.

Finally, recent works by [7] present a lane change intent detection system on highway. They perform detection of rear, side and front vehicles using a combination of different sensory inputs and obtain a reliable scene representation.

M. Garcia Ortiz is with the CoR-Lab, Bielefeld University, Bielefeld, Germany michael.garcia.ortiz@gmail.com

F. Kummert is with the Faculty of Technology, Bielefeld University, Bielefeld, Germany Franz@techfak.uni-bielefeld.de

J. Schmüdderich is with Honda Research Institute Europe GmbH, Offenbach am Main, Germany firstname.lastname@honda-ri.de

978-1-4673-3063-3/12/$31.00 ©2012 IEEE



Fig. 1. Example of an approach of a preceding vehicle followed by a lane change. One image every 2 seconds are displayed, and organized from left to right and top to bottom.

Driver monitoring (head orientation and blinker state) is used in addition to the surrounding monitoring. Using these sensory inputs, they can predict the lane change up to 3 seconds before it occurs. As stated in their contribution, the performance of the system is mainly due to the head monitoring.

These works are of great value to explore the possible capabilities of behavior intent prediction systems. However, they all include monitoring of the driver and rear sensing. We want to evaluate the possibility of driver behavior prediction using only front sensors, limiting as much as we can the amount of sensors. Another reason for avoiding driver monitoring is that using only transposable quantities to build the system (e.g. speed or acceleration) allows the knowledge learned from the point of view of the ego-vehicle driver to be applied to other traffic participants behavior prediction (see, e.g., [17]), which is not possible if part of the system is based on driver monitoring.

III. Methods

An experimental car is equipped with front cameras and front radar. The prediction of the future behavior of the driver is based on the current front scene. Inputs from sensors are filtered and used to compute indicators, which represent relevant information about the current observed scene. We learn the mapping between the current observed indicators and the future behavior of the driver. This approach is compatible with online learning: the system can wait for the occurrence of a behavior to learn the past causes (learning with delay). In this case, we would learn the mapping between the past scene representation and the current behavior.

The behavior of the driver is a binary target: Is the driver going to perform an action in the future prediction horizon? In this sense, we have to solve a classification problem.

We used Locally-Weighted Projection Regression (LWPR) as the learning algorithm. It is fast and suitable for online learning. This algorithm is designed to build incrementally a knowledge representation in the form of receptive fields (internal representation). We believe that the algorithm used is not important. What matters most is the discriminative power of the indicators for the learning task. We used the publicly available implementation of [18] by the authors for all described experiments.

In the training phase, we train the LWPR using the indicators as an input and the labeled future behavior as a learning target. A representation of the system can be observed in Fig. 2.

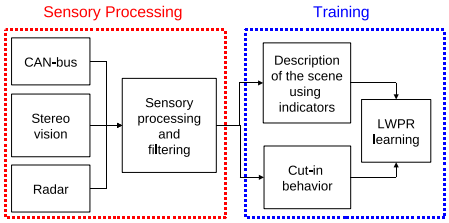

Fig. 2. The mapping between present descriptors of the scene and the future behavior is learned by the LWPR.

In the prediction phase, we use the internal representation of the LWPR to estimate if the driver is going to perform a lane change in the next 2 seconds. We then can compare the prediction to the ground truth and evaluate our prediction system (see Fig. 3).

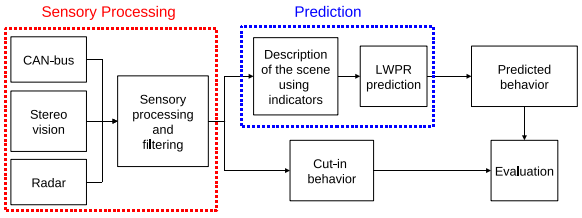

Fig. 3. The mapping learned by the LWPR is used to predict the future behavior for scenes that were not previously encountered.

A. Sensory Inputs

The CAN-bus provides information from which we derive filtered quantities about the ego-vehicle. At each timestamp t_0, we have:

- the longitudinal and lateral absolute velocity: $AbsLongVel_{ego}(t_0)$ and $AbsLatVel_{ego}(t_0)$
- the longitudinal and lateral absolute acceleration: $AbsLongAcc_{ego}(t_0)$ and $AbsLatAcc_{ego}(t_0)$

The sensory system provides information about the surrounding vehicles. We consider only the frontal predecessor

$V_{frontal}$, the left predecessor V_{left} and the right predecessor V_{right}, as described in Fig. 4.

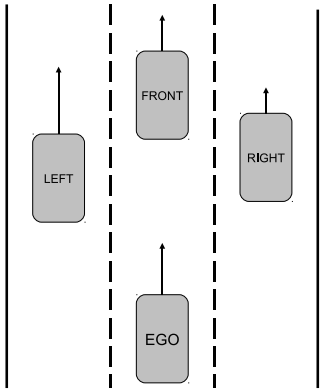

Fig. 4. The sensors detect the frontal successor, the left successor and the right successor as well as the lanes. In most cases, only one or two of these vehicles are present in the scene.

The sensory system provides for each vehicle i at time t_0:

- the relative lateral and longitudinal positions: $LatPos_i(t_0)$ and $LongPos_i(t_0)$
- the lateral absolute and lateral relative velocity: $AbsLatVel_i(t_0)$ and $RelLatVel_i(t_0)$
- the longitudinal absolute and longitudinal relative velocity: $AbsLongVel_i(t_0)$ and $RelLongVel_i(t_0)$
- the longitudinal absolute and longitudinal relative acceleration: $AbsLongAcc_i(t_0)$ and $RelLongAcc_i(t_0)$

B. Indicators

We selected indicators in order to include knowledge in the system: these quantities are descriptors of the current scene. Their level of abstraction is higher than basic sensory inputs. For each timestamp t_0 we compute:

- The time to contact to the frontal predecessor which quantifies the way the driver is approaching the frontal predecessor:

$$TTC_{frontal}(t_0) = \frac{LongPos_{frontal}(t_0)}{RelLongVel_{frontal}(t_0)} \quad (1)$$

- The difference to the desired speed, measured by calculating the average past speed in a time window of δT (5 minutes in this contribution) and comparing it to the speed of the frontal predecessor:

$$\Delta Vel(t_0) = \frac{1}{\delta T}\int_{t_0-\delta T}^{t_0} AbsLongVel_{ego}(t)\,\mathrm{d}t - AbsLongVel_{frontal}(t_0) \quad (2)$$

- The distance to the left predecessor gives information on the availability of the left lane:

$$LongPos_{left}(t_0) \quad (3)$$

- The difference between the left and frontal predecessor speed. If this difference is high, the gain the driver will have by changing lane increases:

$$\Delta Vel_{left-frontal}(t_0) = AbsLongVel_{left}(t_0) - AbsLongVel_{frontal}(t_0) \quad (4)$$

All the indicators are normalized between 0 and 1 in order to approach 1 when the likelihood of lane-change increases (see Fig. 5).

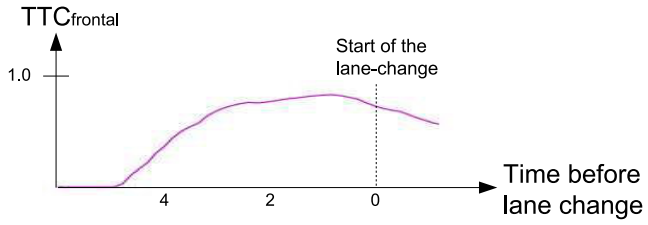

Fig. 5. Evolution of the normalized $TTC_{frontal}$ before a lane change.

C. Target of the prediction system

We want to predict the left lane change, and we define the start of the maneuver by the moment when the car starts turning. We labeled by hand the moment when the car starts turning by observing the video streams. We labeled the frames before the lane change as follows:

- up to 2 seconds before the lane change starts: label 1
- 2 to 4 seconds before the lane change: not used for training
- more than 4 seconds before the lane change: label 0

We do not perform learning from 4 to 2 seconds before the lane change in order to facilitate the task of the learning algorithm, which will identify more easily the difference of features by comparing their distribution before 4 seconds and between 0 to 2 seconds. The temporal window without learning has been chosen empirically. The labeling of a lane change scene can be observed on Fig. 6.

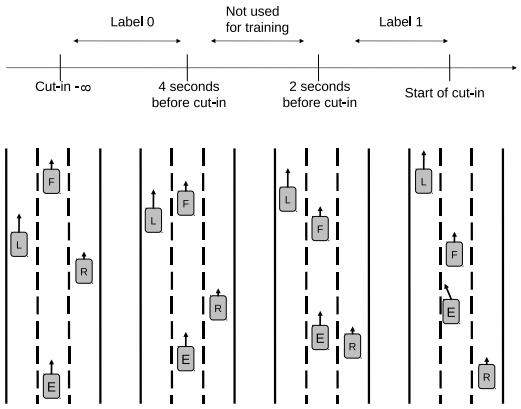

Fig. 6. Labeling of the scene before a lane change event.

D. Building of the datasets

As stated before, one of the aims of this contribution is to evaluate the importance of rear sensing by comparing

prediction of scenes with and without contribution from rear vehicles. For this we created two datasets:

- the first one includes all the previous indicators
- the second one includes all the previous indicators in addition to a label representing the presence of an incoming left successor, simulating rear sensing.

With this two datasets, we can evaluate the importance of rear sensing, and we can test if a prediction system based only on front sensing can be reliable enough.

IV. Evaluation Measures

The evaluation in several contributions about future behavior prediction is performed frame-based: the system evaluates for each timestamp if the prediction is correct or not. This evaluation has a real scientific value, because it is a direct measurement of the learning capability of the system. However, it is not suitable if we want to evaluate the real-world capability of the system. We add to this standard method the event-based evaluation: we group consecutive prediction together, forming temporally consistent events. N successive predictions of a future behavior form an event, and we check if it intersects with a target event. This type of evaluation is really useful to identify clusters or incorrect predictions at the scene level. Also, it enables us to measure the amount of false positive per hour, a measurement commonly used in the automotive industry.

1) N-Fold cross-validation: We use N-Fold cross validation in order to evaluate our system. We split the dataset into N subsets, each containing a scene of the ego-vehicle approaching a frontal predecessor. We train the system using N-1 subsets and we present the samples from the remaining subset to the trained prediction system.

We then use the activations from the N evaluation subsets, obtained from the N possible combinations of training and evaluation subsets, in order to evaluate the quality of the prediction over the whole dataset.

The output of the LWPR gives us a sequence of activations which is normalized by mean and variance. Then, for each timestamp t_0 we obtain a classification value $C^{\text{class}}(t_0)$

2) Frame-based evaluation: The output of the LWPR gives us a classification value for each timestamp $C^{\text{class}}(t_0)$ (simplified C^{class}) We can set a classification threshold τ^{class}, and make a classification decision for each prediction:

$$\begin{aligned} \text{if } C^{\text{class}} > \tau^{class} &: \text{behavior is predicted} \\ \text{else } C^{\text{class}} \leq \tau^{class} &: \text{behavior is not predicted} \end{aligned} \quad (5)$$

For each threshold τ^{class}, we compute the detection rate ν^{correct} and the false positive rate $\nu^{\text{incorrect}}$ which are defined as:

$$\begin{aligned} \nu^{\text{correct}} &= \frac{\#(\text{correct classifications})}{\#(\text{positive examples})} \\ \nu^{\text{incorrect}} &= \frac{\#(\text{incorrect classifications})}{\#(\text{negative examples})} \end{aligned} \quad (6)$$

By varying the classification threshold τ^{class}, a receiver-operator-characteristic (ROC) can be generated. This performance measure is a standard tool in machine learning and has been used to evaluate behavior prediction systems (see, e.g., [14]). In the presented ROC, we plot the detection rate against the false positive rate.

3) Event-based evaluation: For each threshold τ^{class}, we obtain a sequence of decisions for the prediction. For each timestamp, the future lane change is either predicted or not. We group together consecutive positive predictions of the lane change behavior. One interval of these predictions is called an event.

If this event intersects with the 2 seconds interval before the lane change, we consider that the prediction is correct, and we count one true positive event. If this event does not intersect with the 2 seconds interval before the lane change, the prediction is incorrect, and we count one false positive event. We count the amount of of false positive events per hour for each value of the threshold. In particular, we can decide to set the threshold τ^{class} in order to obtain a certain percentage of correctly detected events, and evaluate the amount of false positive per hour for this percentage of correctly detected events. In addition, we calculate the average length of the prediction event.

V. Experimental setup

Our experimental vehicle is equipped with a front camera and a front radar recording data at a frequency of 10 Hz. A combination of radar and video is used to detect traffic participants and their position relative to the ego-vehicle. From this information, we compute their position, speed and acceleration, processed using a Kalman Filter. The cameras are used to detect lane markings, in order to assign a lane to each traffic participant. The sensory setting is described in [19], and illustrated in Fig. 7. The lateral position extracted with radar is refined using video processing.

We use the CAN-bus data in order to obtain information about the driver behavior: the speed and acceleration of the ego-vehicle are also processed using Kalman filtering. We choose not to use blinker status or driver head monitoring, as motivated in II.

The start of the lane change behavior has been annotated. In the future, if we want the system to learn online and adapt to the driver behavior, it can be replaced by a detection of change in lane orientation.

Our data for this study has been acquired by recording the driving of the ego-vehicle on German highway. 50 minutes of driving were recorded under classic driving condition. We labeled 20 examples of lane change situations:

- 10 scenes are normal lane change scenes: the ego-vehicle approaches the frontal predecessor and cuts-in
- 5 scenes are lane changes caused by the right predecessor cutting in in front of the ego-vehicle
- 5 scenes are delayed lane change situations: the ego-vehicle approaches the front predecessor but slows down and waits before cutting in, because of an incoming left successor.

These lane change scenes represent usual situations encountered during driving. As an example, we show one of the normal lane change scenes in Fig. 1

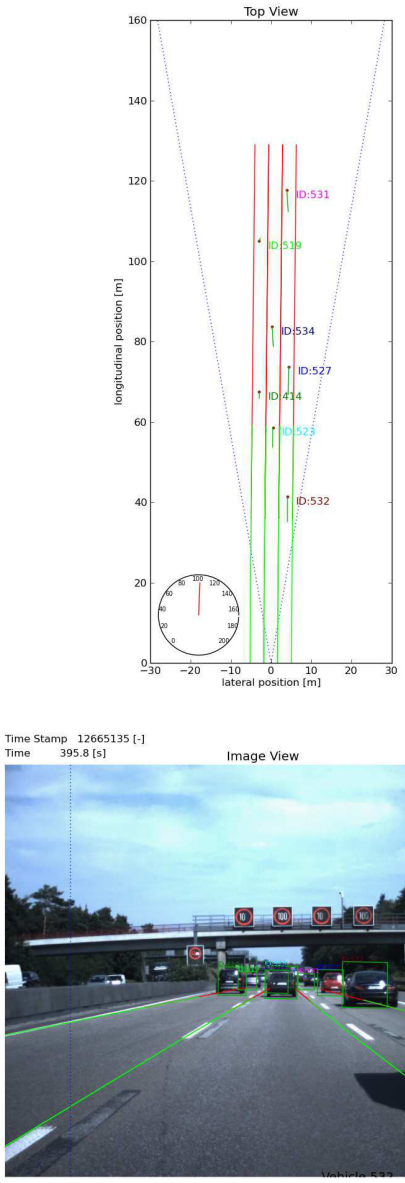

Fig. 7. Result of the sensory processing: vehicle and lane detection illustrated with ego-vehicle perspective and top-view perspective.

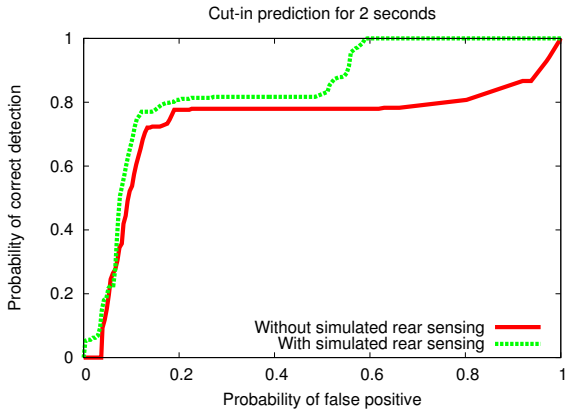

Fig. 8. ROC presenting the results for the frame-based evaluation. As expected, the results are better using a simulated rear sensing.

VI. Experiments and Results

A. Results for future behavior prediction

The correct detection rate against the false detection rate is presented in the ROC curve in Fig. 8. This frame-based evaluation shows that the prediction capability of the system improves slightly if the dataset includes information about rear vehicles. The evaluation also shows that the prediction capability even without relying on driver monitoring is of good quality.

We evaluated the amount of false positive events per hour in the same condition. We set the classification threshold in order to obtain 80 % of correct detected events.

If we use the dataset for which a label representing an incoming left successor was included, we measure 32 false positive events per hour. The average length of the events is 33 timestamps (3.3 seconds) for the positive events, and 37 timestamps (3.7 seconds) for the negative events.

Without this labeling of an incoming left successor we measure 45 false positive events per hour. The average length of the events is 40 timestamps (4 seconds) for the positive events, and 48 timestamps (4.8 seconds) for the negative events. It is important to note that a similar system using only the time to contact $TTC_{frontal}(t_0)$ reaches a performance of 52 false positive events per hour. That means that a huge part of the necessary information is contained in this indicator.

The results demonstrate that the system can learn the future behavior of the driver even without driver monitoring. The prediction without rear sensing is possible, but the results are not as good as with simulated rear sensing.

B. Characterization of the errors

Using event-base evaluation, we could identify several factors responsible for false positives that can be taken into account in future experiments. First of all, multiple detection errors occurred during the processing of the sensory inputs, and were responsible for 5 false positive events. We will have to explore the possible improvements for the sensory setting. The detection can be improved by having a better fusion of information. A second possible reason for false positive is the case when the driver is in a traffic jam, or a congested traffic. 5 false positive events occurred during traffic jams: the driver does not change lane, even if he can, because there is no gain for him. A third reason is the case when the driver wants to exit the highway in the near future (3 false positive events). He stays on the right lane, and again, even if he has the occasion to change lane, and even if in normal conditions he would have changed lane, he does not.

VII. Discussion

We presented a set of indicators which are used to perform behavior prediction on highway. The system predicts if the ego-vehicle driver is going to start a change to the left lane in the next 2 seconds. This prediction does not require

any driver monitoring. We evaluated the importance of rear sensing, and showed that if the prediction was not completely dependent on rear sensing, knowledge about the rear scene improves the results of the prediction.

Considering the labeling of the behavior of the driver, we believe that this label can be replaced if we want to implement the system online by measuring the car orientation compared to the lane. The start of the lane change would then be defined at the moment when the orientation of the car relative to the lane reaches a value higher than a certain threshold, or when the car crosses the lane. A simple way of labeling the driver behavior would be to define the start of the intent of lane change as the moment when the driver starts using his blinker. However, it highly depends on the driving habits of the driver: some drivers use their blinkers far in advance, and other use them at the last moment, or even do not use them at all.

The results are promising, but the prediction is not reliable enough to be used in a commercial product as is. The number of false positive events per hour is low in absolute value, however, one has to consider human tolerance for errors: a driver will probably turn off a system which is wrong too often. We believe it is possible to improve the system, based on the identification of the main reasons for false positive events (see Sec. VI-B). Compared to state-of-the-art approaches our system shows convincing performance. However, for several reasons, we consider a direct prediction to be unsuited: the dataset is different, the drivers are different, and finally the countries where the recording occurred are most of the time different. These factors might influence greatly the predictability of the behavior.

VIII. Future Works

First of all, we are planning to record more driving scenes on highway. We want to show that our approach generalizes well, so we need longer recordings. The data used for this contribution were recorded with two different drivers. However, the quantity of available data did not allow us to compare the influence of driving styles properly. So we want to record longer driving sessions with different drivers, in order to evaluate the adaptation capability of our system, as well as the influence of the driving style of the driver.

One important extension of the system would be to develop an indicator giving information about the occupancy of the highway. This indicator could tell the system if we are in a dense traffic, or in a traffic jam. We believe that the system will create different behavior models depending on the value of this new indicator.

When a driver approaches a preceding vehicle, he can either change lane (left or right) or stay in the lane but reduce his speed. A necessary improvement would be to predict both the lane change behavior and the braking behavior in order to introduce competition between the classification outputs. This affordance mechanism is observable in the human brain, and we think that it could lead to better results.

Finally, an important extension would be to perform the prediction at multiple time-scales and identify the time of prediction leading to the best results. Also, this prediction at multiple timescales can be used as a temporal integration, or a verification for temporal consistency, in order to remove false positives.

References

[1] F. Moosmann and C. Stiller, "Velodyne slam," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2011.

[2] J. Schmüdderich, N. Einecke, S. Hasler, A. Gepperth, B. Bolder, R. Kastner, M. Franzius, S. Rebhan, B. Dittes, H. Wersing, J. Eggert, J. Fritsch, and C. Goerick, "System approach for multi-purpose representations of traffic scene elements," in *Proc. of the IEEE Intelligent Transportation Systems Conference*, 2010.

[3] R. Kastner, T. Kühnl, J. Fritsch, and C. Goerick, "Detection and motion estimation of moving objects based on 3d-warping," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2011.

[4] X. Ma, "A neural-fuzzy framework for modeling car-following behavior," in *Proc. of the IEEE International Conference on Systems, Man and Cybernetics*, 2006, pp. 1178 – 1183.

[5] T. Hülnhagen, I. Dengler, A. Tamke, T. Dang, and G. Breuel, "Maneuver recognition using probabilistic finite-state machines and fuzzy logic," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2010, pp. 65 – 70.

[6] Y. Liu, H. Zhang, H. Meng, and X. Wang, "Automatic mining of vehicle behaviors with an unknown number of categories," in *Proc. of the IEEE Intelligent Transportation Systems Conference*, 2008, pp. 247 – 252.

[7] B. Morris, A. Doshi, and M. Trivedi, "Lane change intent prediction for driver assistance: On-road design and evaluation," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2011.

[8] M. Garcia Ortiz, J. Fritsch, F. Kummert, and A. Gepperth, "Behavior prediction at multiple time-scales in inner-city scenarios," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2011.

[9] T. Gindele, S. Brechtel, and R. Dillmann, "A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments," in *Proc. of the IEEE Intelligent Transportation Systems Conference*, 2010.

[10] M. Reichel, M. Botsch, R. Rauschecker, K. Siedersberger, and M. Maurer, "Situation aspect modelling and classification using the scenario based random forest algorithm for convoy merging situations," in *Proc. of the IEEE Intelligent Transportation Systems Conference*, 2010, pp. 360 – 366.

[11] A. Gerdes, "Automatic maneuver recognition in the automobile: the fusion of uncertain sensor values using bayesian models," in *Proc. of the 3rd International Workshop on Intelligent Transportation*, 2006.

[12] I. Dagli, M. Brost, and G. Breuel, "Action recognition and prediction for driver assistance systems using dynamic belief networks," in *Agent Technologies, Infrastructures, Tools, and Applications for E-Services*, ser. Lecture Notes in Computer Science, 2003.

[13] M. Heracles, F. Martinelli, and J. Fritsch, "Vision-based behavior prediction in urban traffic environments by scene categorization," in *Proc. of the British Machine Vision Conference*, 2010.

[14] J. C. McCall, D. Wipf, M. M. Trivedi, and B. Rao, "Lane change intent analysis using robust operators and sparse bayesian learning," in *IEEE Trans. on Intelligent Transportation Systems*, 2005, pp. 431–440.

[15] M. J. Hennings, O. Georgeon, T. Wynn, and J. F. Krems, "Modelling driver behavior in order to infer the intention to change lanes," in *Proc of European Conference on Human Centered Design for Intelligent Transport Systems*, 2008.

[16] N. Oliver and A. Pentland, "Driver behavior recognition and prediction in a smartcar," in *Proc. of SPIE Enhanced and Synthetic Vision*, 2000.

[17] M. Garcia Ortiz, J. Schmüdderich, F. Kummert, and A. Gepperth, "Situation-specific learning for ego-vehicle behavior prediction systems," in *Proc. of the IEEE Intelligent Transportation Systems Conference*, 2011.

[18] S. Vijayakumar, A. D'Souza, and S. Schaal, "Incremental online learning in high dimensions," in *Neural Computation*, 2005.

[19] M. Nishigaki, S. Rebhan, and N. Einecke, "Vision-based lateral position improvement of radar detections," in *submitted to the Intelligent Transportation Systems Conference*, 2012.