Proceedings of the 11th International IEEE
Conference on Intelligent Transportation Systems
Beijing, China, October 12-15, 2008

Continuous Driver Intention Recognition with Hidden Markov Models

Holger Berndt, Jörg Emmert, and Klaus Dietmayer
Institute of Measurement, Control, and Microtechnology
Ulm University, D-89081 Ulm, Germany
Tel: +49 731 50 26329, Fax: +49 731 50 26301
Email: holger.berndt@uni-ulm.de

Abstract—**The most common cause of accidents in individual road traffic is human failure. Accidents often arise from misbehavior of one or several drivers when inducing a driving manoeuvre. Dangers can occur either when the intented manoeuvre is not well adjusted to the current traffic situation, or when the manoeuvre is not properly announced to the environment so that the intention is misinterpreted. When designing advanced driver assistance systems, it is beneficial to gather information about driver behaviors as accurately and early as possible. This work investigates early driver intention inference with Hidden Markov Models by observing easily accessible vehicle and environment signals such as pedal positions or global vehicle position on a digital map in real traffic.**

I. Introduction

Driving a car has nowadays become an increasingly complex task. On the one hand, rising traffic density demands increased attention from the driver, on the other hand, the number of sources capable of distracting the driver, such as cellular phones or navigational devices, is also on the rise. At the same time, accident statistics have shown that the vast majority of accidents are caused by the drivers' operating errors. For example, German statistics counted 1.3 human misbehaviors for each accident in 2005, comprising more than half of all involved persons [1]. For this reason, it seems purposeful to prominently include the main factor of danger, the human behavior, in the design of advanced safety systems. In particular, safety systems that aim to warn the driver, and therefore need to have a rather long temporal horizon, benefit from knowing which manoeuvre the driver of the host vehicle is performing at any given time; or better yet, which manoeuvre is likely to be imminent. Safety systems that automatically intervene in the vehicle dynamics and therefore share vehicle control with the human driver, can also take advantage of the driver's expected behavior to ensure that both parties are not operating against one another, see Fig. 1.

Manoeuvres can be seen as an implementation of the driver's intention on the tactical level of the scheme introduced by Michon [2]. As the intention is an inner state of the driver and thus cannot be observed directly, it has to be infered from observable signals from the driver and the environment in which the car operates. In general, many signals that describe how the driver influences the vehicle dynamics are appropriate for this observation, for example the steering wheel angle and velocity, pedal positions, and turn indicator activations. As these signals are usually available on the CAN-bus of the vehicle, they can be obtained relatively cheaply. It is also possible to observe the driver with a video camera, but as this is more expensive and may involve acceptance issues from drivers, those signals are not used in this work. Other signals may describe the environment as well as the car's relation to the environment. Examples for those signals may be the lateral position of the car in the lane or street information from a digital map of a navigational device.

Information about a forthcomming manoeuvre can be used to predict a more accurate trajectory of the host vehicle, which in turn can lead to a more reliable collision risk assessment [3]. Knowledge of the driver's intention can also be used to minimise unwanted interference between mutually exclusive driver assistant systems, such as a lane keep and a lane change assistant: If for example a driver's intention to overtake another car has been recognised, a lane change assistant can be switched on, otherwise, the lane keep assistant may be active.

It is not sufficient to rely exclusively on indicators such as the turn signal for intention recognition, because although legally mandatory in many countries, drivers tend not to use the signals consistently. Olsen showed in [4] that the turn indicator is only used in 66% of the lane changes, while only about 50% of all turn indicator activations occur during the initial phase of the lane change manoeuvre [5]. Moreover, situations in which the driver fails to appropriately announce the manoeuvre to the environment via the indicators may be the most dangerous and therefore interesting ones.

Several attempts have been made to recognise a specific driver intention during the development of assistant systems, such as a predictive braking assistant [6]. Since driver intention recognition is basically a classification problem, most published systems are based on well known techniques for pattern recognition, such as Support Vector Machines [7]. Other techniques try to emphasise the psychological aspects of the human car operator by embedding driver models into cognitive frameworks [8]. The ability to natively represent time variations in the observable signals as well as to cope with remarkable driver- and situation-dependent variations have made Hidden Markov Models (HMM) [9] a popular methodology in the area. The general idea behind the use of HMM is that a driver can be modeled as having a (possibly large) number of internal mental states. Progressing through those internal mental states during driving, the driver develops the intention to start a manoeuvre. In turn, a manoeuvre can

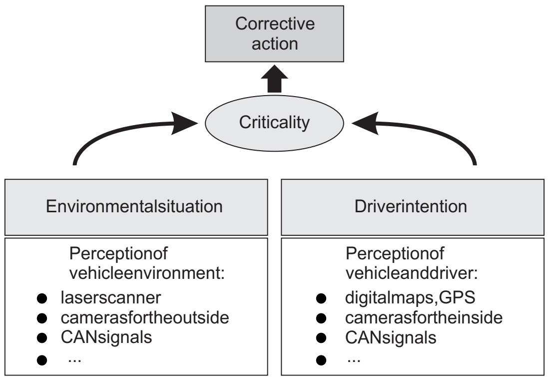


Fig. 1. Placement of driver intention recognition for driver assistant systems. In order to have a better chance of assessing the criticality of a situation and providing the best possible human-machine interaction once a danger has been identified, non-autonomous assistance systems need to have an idea of the driver's current intention in addition to the model of vehicle surroundings.

be broken up into a chain of consecutive actions, which can be modeled and analysed as a HMM [10], [11].

Driver intention recognition can be done in a distcrete or continuous manner. In a discrete recogniser, each manoeuvre is modelled as a Markov model, which has been trained by a set of sample data for the complete manoeuvre. Such a system, when presented with a time window of batched signals from the last few seconds, has the potential to recognise the whole manoeuvre after it has been completed. This is similar to the use of HMM in other disciplines, such as speech or handwriting recognition, where words or text should be identified after they have been spoken or written. For driver assistant systems, knowing a manoeure after it has been completed may be useful to classify the current driver as sporty or defensive. However, for advanced driver assistant systems, it is much more interesting to identify a manoeuvre at a very early stage after it was started. This can be accomplished with continuous recognition, where a fixed-length time window of signals is batched and continuously tested against the Markov models [8].

Even though previous tests have shown promising potential in the use of HMMs in driver intention recognition systems, many aspects, such as the optimal HMM structure and grammar for each individual maneuvre, or the most effective input signals for a respective manoeuvre, are still to be investigated. Tests have shown that it is not ideal to use all available input signals for the recognition of all manoeuvres, but that it is important to limit the signals to the most representative ones for each manoeuvre individually. This work aims to give a contribution in this field.

II. System Setup

The system operates with what can be regarded as standard sensors in modern cars. Information about velocity, yaw rate, pedal positions, steering angle, steering angle velocity, and turn indicator signal activations are nowadays easily accessible over the CAN-bus without the need of extra sensors. Also, with ever rising market penetration of navigational devices, a GPS sensor and a digital map can be expected to be present in future cars.

Features for the HMM are generated by applying a series of operations to the observed signals. First, all signals are batched for a predefined fixed-length time interval and equalised to the same frequency. Also, a number of operators can be applied to the signals. Signals from noisy sensors can be smoothened by suitable temporal filters or new features can be generated by combining different observed signals or applying e.g. a differential operator.

These features can then be used as inputs to the HMMs at the matching stage, as well as in pre- or postprocessing rules for the recogniser, see Fig. 2. It is even possible not to test some of the models if certain conditions are not met in the preprocessing stage, or to modify the likelihood of certain manoeuvres as part of a plausability check on post-processing.

Continuous signals, such as the steering wheel angle, are predestinated to act as features for the HMMs. Signals where only the current value, rather than the progress, is significant, are referred to as status signals. Those signals often appear in binary form, such as the turn indicators, and are well suited to be used in a postprocessing stage to alter the likelihoods towards manoeuvres that fit the current signal value better.

The position on a digital map can be used as an additional sensor from which various pieces of information can be extracted. Most apparently, discrete status messages can be extracted. For intention estimation, it is for example interesting to know how many lanes are on the road, the road type, the speed limit and whether any changes in the speed limit have recently occured. As the environment renders some manoeuvres impossible (e.g. turning left on a highway), those do not even have to be tested, thus saving computational cost and reducing the classificator dimension in these scenarios.

It is also possible to extract continuous signals from a digital map. The likelihoods of a left or right turn manoeuvre to be started at any given point in time can be expected to be correlated with the distance to the next crossing where such a manoeuvre is possible and/or permitted.

A very interesting application of digital maps is the improvement of the signal quality of vehicle signals by putting them into environmental context. Steering wheel activity may start a lane change or turn manoeuvre, but it may also just represent a lane keep manoeuvre on a curvy road. For intention recognition, a more explicit signal than the raw steering wheel angle is the deviation of the current steering wheel angle from the angle required for normal lane following. By extracting the street curvature from the digital map and using the steering wheel transmission to calculate back to the steering wheel angle needed for lane following, a more meaningful feature can be generated. Fig. 3 shows a plot of the steering angle received from the CAN-bus during normal street following and the calculated steering angle that is necessary to follow the curvature of the road. Although the street curvature extracted from the self-generated digital maps in this work is only a very rough estimate and will be improved in future work, it can be seen from Fig. 3 that the signals behave uniformly,

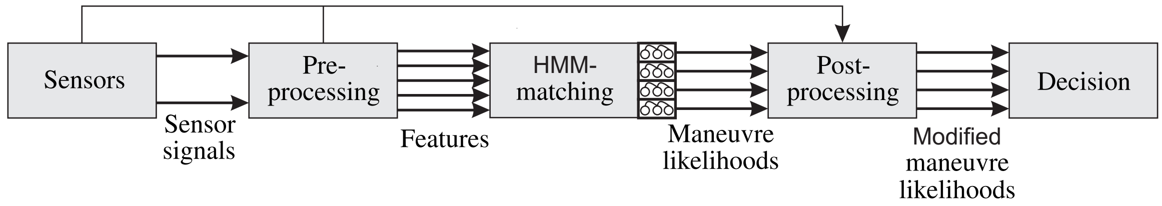


Fig. 2. Signal flow of the recognition process. Sensor signals can be pre-processed and combined to form features for the HMM recognition process. The resulting manoeuvre likelihoods can be modified in a post-processing stage, based on raw sensor or preprocessed signals for a final decision of the most likely manoeuvre.

and although the differential signal will not vanish, it will be less biased. In future works, current street curvature data might be taken from online driving tube estimations which can be performed on online maps [12]. That approach would also minimise the influence of the position uncertainty on the current street curvature estimate.

As drivers' intentions are often heavily influenced by surrounding traffic, it seems feasible to include information about adjacent vehicles in the estimation process. Although the test vehicle is equipped with an IBEO AS laser scanner, and tracking [13] and object classification [14] algorithms are available, they were not included during the estimation phase. Despite our belief that the recognition results would have been improved, the risk to overrate 'sane' driving manoeuvres and fail just in those situations where a real danger evolves seems too high. For the same reason, restrictions introduced based on map and turn signal information must be handled with care. A manoeuvre at an upcomming intersection might very well be intended despite being legally forbidden.

III. System Training

A statistical model, such as a Hidden Markov Model, has to be trained with labeled sample data to determine the model parameters in an optimal fashion before it can be used for intention recognition, see Fig. 4. Based on the training set, the *Baum-Welch* algorithm can be used for parameter estimation. The implementations of the HMM standard algorithms of the Hidden Markov Toolkit (HTK) [15] have been used in this work. It is important to note that because HMMs are used, the segmentation of the data is done automatically during the training process. It is not necessary to manually segment the data and define the invididual phases of the manoeuvre that correspond to the discrete states in the Markov chain.

Collecting training data is an especially cumbersome process, because in contrast to for example object classification by image processing, where oftentimes several samples per second can be extracted from a video stream, driving manoeuvres happen on a much larger time scale. Also, manoeuvres are often combined or aborted which makes collection of clean training data more difficult. In this work, it was decided to use normal test drives not especially created for intention recognition to extract a training set semi-automatically. In order to make it easily possible to compare different features, the feature generation must be separated from the training set labeling. This can be accomplished by performing driving tests with all available sensors active, and then labelling the manoeuvres by specifying manoeuvre identifier together with start and end timestamps. The HMM input generation for a specific feature set can then be done automatically on this database.

The labeling itself can also be algorithmically supported. The idea is that manoeuvres are regarded to be characterised by a deviation of the car from normal lane following. Signals during lane following tend to be smooth and steady, while at manoeuvre start or end, signal peaks occur far more often. This characteristic can be exploited by building finite differences on the continuous signals and comparing them to a threshold. In the same way, switches in binary signals, such as turn indicators, can be searched. If all time intervals are united

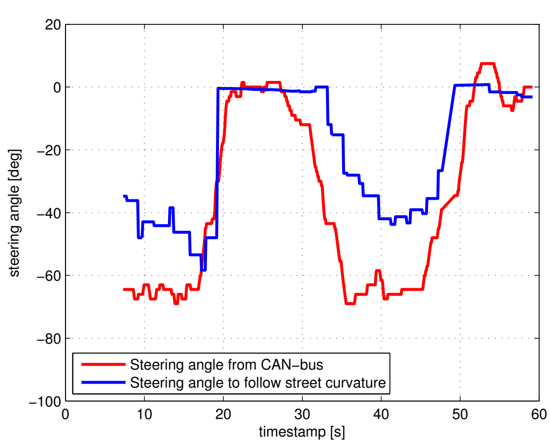


Fig. 3. Steering angle from onboard sensors (e.g. CAN-bus) compared to the estimated steering angle that is needed to follow the road curvature, as extracted from a digital map. The differential signal is less biased and represents the steering angle deviant from normal lane following better.

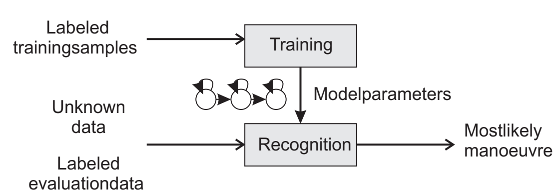
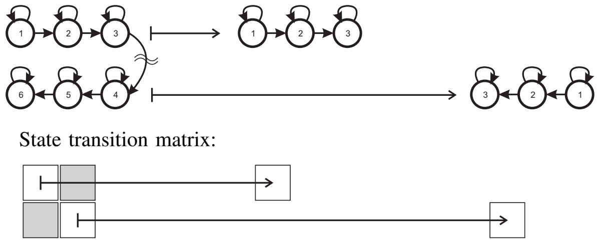


Fig. 4. System overview. The model parameters are estimated from labeled training data. During recognition stage, the parameterised model can be used with unknown data for recognition or with labeled data that was not used for training purposes for system evaluation runs.



(a) Original model (b) Submodel 1 (c) Submodel 2

Fig. 5. Dissolving of model grammar. In this schematic example, a left-to-right Hidden Markov Model consisting of six states is broken up into two models with three states each. The corresponding state transition matrix is likewise broken up into a block matrix, of which only the diagonal parts are preserved. The gray parts of the matrix, which are zero for left-to-right models, are disregarded in the general case.

where at least one of these criteria is active, this set can be regarded as a proposed list of manoeuvres for a given data stream. A human supervisor is still required to check the results, adjust the timestamps and label the identifiers.

IV. Intention Recognition

Intention recognition can be formulated as the evaluation problem for Hidden Markov Models: Given an observation sequence and a set of parameterised models (one for each discrete manoeuvre that is to be distinguished), what is the probability that this sequence has been generated by the respective model? This problem can be efficiently solved with the *forward* algorithm [9].

As pointed out above, recognising manoeuvres only after they have been completed is not desired for this application. The obvious solution to this problem would be to define and train Markov chains that only correspond to the initial phases of a manoeuvre instead of the complete time period. While it is difficult already to define start and end timestamps for complete manoeuvres, this is error-prone for fuzzily defined phases.

Another solution is to train the complete models first, and then dissolve the grammar of the Markov chain, thus effectively extracting submodels from the original model, each corresponding to a short phase during the manoeuvre. Here, the early phases are especially interesting for intention recognition.

The question arises whether the submodels can still represent manoeuvre phases. Mathematically, this procedure effectively results in neglecting far off-diagonal elements of the state transition matrix of the HMMs, Fig. 5. Whether or not the submodels are still parametrised in a meaningful way depends on the influence of the neglected part on the model.

Driving manoeuvres consist of a series of consecutive actions that are characteristic for the manoeuvre. Therefore, an adequate representation of a complete driving manoeuvre is a left-to-right Markov chain.

For models that are not overly coupled, the neglected parts of the state transition matrix are small or zero, so that the resulting submodel grammar structure does not diverge substantially from the original model. In particular for left-to-right models as used in this work, the parameters obtained by training on the full model can be taken verbatim to define the submodels.

Fig. 6 depicts a schematic recogniser which distinguishes two discrete manoeuvres. The model can now be broken into several submodels, which can have overlapping states, called *recombination* in the figure. Of course, the width of the batched signal window has to be adapted to the expected average exposure time in the respective model which can be gained from the transition probabilities. Therefore not all submodels will have the same input vector size.

As the assistance system is only interested in an early intention recognition, only submodels in the early manoeuvre phases may be considered. As the batched time window gets more narrow, false classifications between similar manoeuvres (e.g. turn left and lane change left) become more likely. However, environment dependent context information (e.g. from digital maps) at the pre- or postprocessing stage helps to yield a good overall classification performance at an early stage in time.

V. System Evaluation

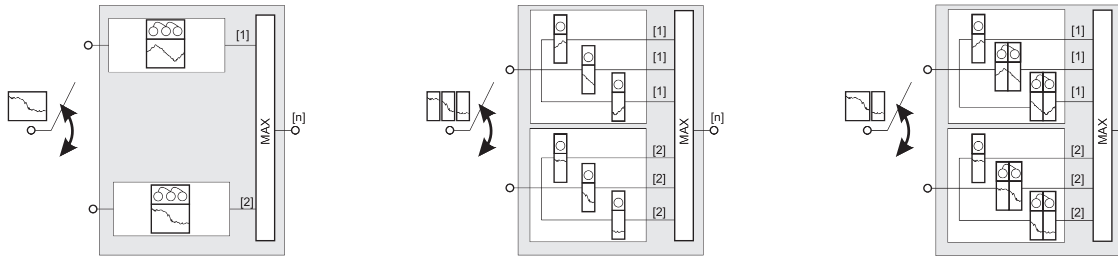
The task of the trained intention recognition system was to distinguish left and right lane change as well as left and right turn manoeuvres from normal lane following. The available input signals were throttle pedal position, brake pressure, cross acceleration, steering wheel angle, steering wheel angle velocity, yaw rate, and velocity from the CAN bus as well as distances to next left and right turn possibility, street curvature, and street type from a digital map. Turn indicators were not used, as the risk of over-weighting them in well-behaved situations and then failing in critical driving situations was estimated to be too high. For each manoeuvre, several combinations of signals, HMM grammars and submodel configurations have been tested.

It was decided that it is better to let the probants drive naturally instead of giving manoeuvre commands which pose the danger of falsifying natural behavior. Unfortunately, this makes it difficult to estimate how early after manoeuvre start the recognition took place, because the manoeuvre start timestamp is not clearly defined. This time delay, however, can be roughly estimated by the average duration of stay in the model from state 1 to the last state used for recognition, which can be expected to be within the first two to five seconds of the manoeuvre.

A. Lane change manoeuvres

The training data set for lane change manoeuvres consisted of about 100 lane changes, 50 for left and right lane change respectively. The evaluation data set consisted of a separate set of various driving sequences with about 50 lane change manoeuvres embedded in about an hour driving, mainly on freeways.

As it was expected that the steering wheel angle is the most dominant feature of the available signals, it has been taken



(a) Discrete recognition

(b) Continuous: Dissolved grammar

(c) Continuous: Recombinated grammar

Fig. 6. HMM grammar structure for discrete and continuous recognition. Models can be split into submodels, which can then be recombined to form new grammars. The length of the input signal window has to be adapted to fit the respective submodel.

separately to gather experience of suitable model grammars (that is, the number of states in the left-right Markov chain) and submodel dimensions. Among the tested grammars were six, nine, and twelve state models. For each model, various submodels have been tested (e.g. first two states, first three states, states two to four etc). The best model was found to be a submodel of the first three states of a trained Markov chain with nine states, see Fig. 7. These corresponded to a window length of the batched signal of slightly below 2 seconds. A whole manoeuvre takes on average about eight to ten seconds, so the detection is possible rather early in the manoeuvre.

In a next step, it was investigated whether using additional signals besides the steering wheel angle improves recognition performance. It was found that, not taking surrounding vehicles and other dynamic context into account, many signals from the CAN bus, like acceleration pedal position or brake pressure, were not representative for a lane change manoeuvre. A car might, for example, depending on the surrounding traffic, either break or accelerate during the lane change manoeuvre. The signals steering wheel angle, steering wheel angle velocity and yaw rate were examined in more detail. All one, two, and three-element tuples of those signals were investigated, resulting in seven combinations. Fig. 8 shows examples of steering wheel angle and yaw rate signals during a lane change manoeuvre. It can be seen that the better signal-to-noise-ratio of the steering wheel angle makes it the dominant signal in this case. The best recognition performance was reached with a model that was trained with the steering wheel angle and steering wheel angle velocity as observation signals.

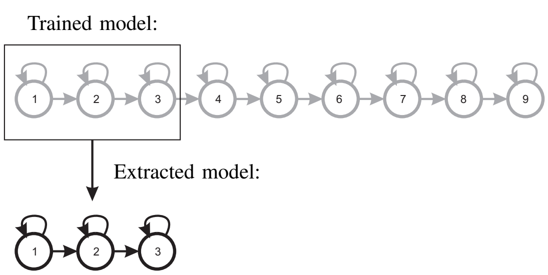


Fig. 7. The best model for lane change manoeuvres consisted of nine states in a left-to-right arrangement. For continuous recognition, a submodel of the first three states was extracted.

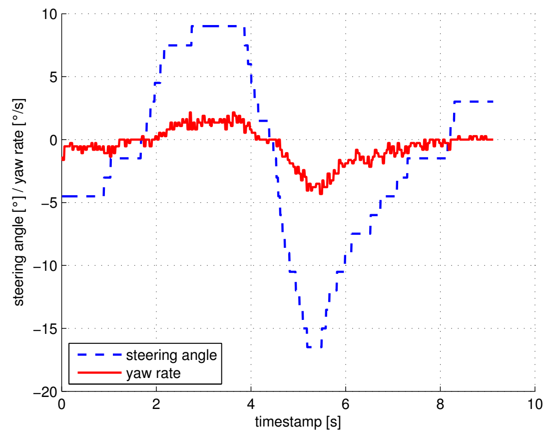


Fig. 8. Steering wheel angle and yaw rate during a lane change manoeuvre. The steering wheel angle is the dominant signal in this comparison.

TABLE I
CONFUSION MATRIX OF LANE CHANGE MANOEUVRE RECOGNITION

		actual manoeuvre		
		LCL	LCR	none
recogniser	LCL	17	0	5
	LCR	0	20	7
	none	7	7	-

As can be read from the confusion matrix in Tab. I, the sensitivity of a detection of a left and right lane change manoeuvre is 71% and 74%, respectively. All lane changes have been detected before the lane markings were crossed.

Taking only manoeuvres on straight roads into account, the sensitivity of the classificator is significantly better. Most miss- or false classifications occur on curvy roads, especially when the lane change direction is opposed to the road curvature. Once our map-based road curvature estimation will be improved (Fig.3), we expect better results on curvy roads as well. To investigate this effect quantitatively, however, a bigger set of test data will be required.

TABLE II
CONFUSION MATRIX OF TURN MANOEUVRE RECOGNITION

		actual manoeuvre		
		TL	TR	none
recogniser	TL	15	0	0
	TR	0	20	2
	none	0	0	-

B. *Turn manoeuvres*

The training data set for turn manoeuvres consisted of about 60 turns, 30 for left and right turns respectively. The evaluation data set contained 35 turn manoeuvres embedded in several hours of driving, mainly in suburban areas.

Again, various Markov model grammars have been compared with the steering angle as the only feature. The best model was found to be a trained Markov chain with nine states, from which a submodel consisting of the third until the sixth state were extracted. It is remarkable that the steering into the curve for turning oftentime started very slowly, so this submodel does not contain the the first two states of the full chain. Unfortunately, this also results in a later recognition of the manoeuvre. The batched time window consists of signals of about 4 seconds.

Following these investigations, different signal combinations were researched. The amount of tests done was comparable to the lane change manoeuvres. It was found that the diversity of turn manoeuvres is so vast that none of the additional signals could improve the recognition performance. Many maoeuvres in the database were driven in hilly terrain, so for example acceleration and braking behaviors differ significantly depending on from which side of the intersection the vehicle is approaching.

During the recognition stage, the classificator was restricted to only output turn manoeuvres within 30 meters of a crossing, based on digital maps. The confusion matrix is shown in Tab. II.

The sensitivity of the classificator shows good results with for the evaluation set perfect detection rate. The two missclassifications were induced by two evasion manoeuvres around parked cars in close neighborhood to a crossing. However, the data set for the turn manoeuvres also clearly needs to be expanded.

VI. CONCLUSIONS AND FUTURE WORK

We have presented a first approach of a framework for early intention recognition of human drivers based solely on easily accessible signals in modern cars. Initial tests showed promising recognition results. The system is able to recognize a set of driving manoeuvres at an early stage even if only a limited set of input signals are used.

While we believe that this system has a lot of potential, a lot more work is needed in order to make the classifier more robust. We are currently working on making more context information available to the system. In particular, with lane-keep assistants getting more and more popular, the lateral position of the vehicle in the lane, as well as the angle of the vehicle towards the lane can be considered to be available signals as well. It is believed that this information would help significantly during intention recognition, especially of lane change manoeuvres.

It is also likely that calculating new features from correlated signals would be helpful. While e.g. the velocity, throttle pedal, and brake pressure themselves may not be representative signals for a turn manoeuvre, a combination of those, maybe taking a height profile from a digital map into account, may effectively describe the manoeuvre.

Also, we're continuously increasing the size of the manoeuvre database. More training and evaluation data is needed for further testing. Besides that, up to now, the data collected was without exception derived from driving tests of proficient male drivers of ages between 20 and 30 years. This is not a completely representative sample of average drivers and more elaborate system testing is planned to be done in future works.

VII. ACKNOWLEDGMENTS

This work has been supported by a grant from the Ministry of Science, Research, and the Arts of Baden-Württemberg, Germany (Az: 23-7532.24-14-19).

REFERENCES

[1] Statistisches Bundesamt, *Verkehr – Verkehrsunfälle 2005*, Fachserie 8, Reihe 7 ed., Wiesbaden, Germany, 2006.
[2] J. A. Michon, *Human Behaviour and Traffic Safety*. Plenum Press, 1985, ch. A Critical View of Driver Behaviour Models: What Do We Know, What Should We Do?, pp. 485–520.
[3] H. Berndt, S. Wender, and K. C. J. Dietmayer, "Driver braking behavior during intersection approaches and implications for warning strategies for driver assistant systems," in *IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, June 2007.
[4] E. C. B. Olson, "Modeling slow lead vehicle lane changing," Ph.D. dissertation, Industrial and Systems Engineering, Virginia Polytechnic Institute, 2003.
[5] S. E. Lee, E. C. B. Olson, and W. W. Wierwille, "A comprehensvie examination of naturalistic lange changes," National Highway Traffic Safety Administration, U.S. Department of Transportation, Tech. Rep., 2004.
[6] J. C. McCall and M. M. Trivedi, "Human behavior based predictive brake assistance," in *IEEE Intelligent Vehicles Symposium*, June 2006, pp. 8–12.
[7] H. M. Mandalia, "Driver intentions," Drexel University, Tech. Rep., 2004.
[8] N. Kuge, T. Yamamura, O. Shimoyama, and A. Liu, "A driver behavior recognition method based on a driver model framework," in *SAE World Congress*, Detroit, USA, March 2000.
[9] L. R. Rabiner and B. H. Juang, "An introduction to hidden markov models," *IEEE ASSP Magazine*, vol. 3, pp. 4–16, 1986.
[10] A. Pentland and A. Liu, "Modeling and prediction of human behavior," *Neural Computation*, vol. 11, no. 1, pp. 229–242, 1999.
[11] A. Liu and A. Pentland, "Towards real-time recognition of driver intentions," in *IEEE Intelligent Transportation Systems Conference*, Boston, USA, November 1997.
[12] T. Weiss, B. Schiele, and K. Dietmayer, "Robust driving path detection in urban and highway scenarios using a laser scanner and online occupancy grids," in *IEEE Intelligent Vehicles Symposium*, 2007.
[13] N. Kämpchen, M. Bühler, and K. C. J. Dietmayer, "Feature-level fusion for free-form object tracking using laserscanner and video," in *IEEE Intelligent Vehicles Symposium*, Las Vegas, USA, June 2005.
[14] S. Wender and K. Dietmayer, "An adaptable object classification framework," in *IEEE Intelligent Vehicles Symposium*, 2006.
[15] S. Young, "The HTK hidden markov model toolkit: Design and philosophy," Department of Engineering, Cambridge University, UK, Tech. Rep., 1993.