in: Proc. 15th Int. IEEE Conf. on Intelligent Transportation Systems (ITSC 2012), Anchorage, pp. 888-893, IEEE 2012

Deciding What to Inspect First: Incremental Situation Assessment Based on Information Gain

Matthias Platho and Julian Eggert

Abstract— **In order to offer even more sophisticated functionality, future driver assistance systems need the ability to robustly recognize and understand driving situations. Especially in inner-city scenarios the high complexity and variability of situations encountered make their assessment a challenging task. We propose to tackle these challenges by decomposing situations into smaller, more manageable parts. We define such a part as a set consisting of a road user and all entities (e.g. cars, traffic lights) currently affecting its behavior. Though the decomposition alleviates the assessment already, for higher numbers of present entities the recognition of interrelated entities is still computationally expensive if performed in a brute-force fashion. Therefore we employ sensitivity analysis on Bayesian Networks for sensibly controlling the recognition process on the basis of information gain. This leads to an active measurement process in which a situation is perceived incrementally, concentrating first on the most meaningful sensor measurements. The proposed method is evaluated on a simulated inner-city scenario where it reliably recognizes the affecting entities of each road user. We show that a recognition process based on information gain can save more than 50% of measurements without significantly impairing the recognition rate.**

I. INTRODUCTION

Advanced driver assistance systems are getting increasingly more powerful and thus require an increasingly comprehensive understanding of the current situation. In recent years, research in related fields has gained more and more interest.

The works published so far can be coarsely divided into those that target the assessment of a situation as a whole and those that focus on specific driving situations or maneuvers. The approaches presented by [1] and [2] belong to the former category. In [1] situations are classified by employing case-based reasoning on a set of predefined situation prototypes. Encountered situations are subsequently added to the predefined prototypes, which tackles the usually high variability of situations but is also prone to harming the stability of the classification in the long run. In [2] the authors use description logic for reasoning about the relations between cars in an intersection scenario. While they report to successfully infer relations and possible conflicts between road users, they also mention the high computational effort required for reasoning.

Significantly more works have been published in the second category, the recognition or prediction of specific driving situations. In [3] the current driving situation is assessed in order to judge the possibility for performing a lane change. The authors employ a Bayesian Network to enter evidence about street parameters and surrounding road users and obtain a recommended action. In [4] maneuvers comprising two cars like 'passing' and 'following' are recognized using a combination of Dynamic Bayesian Networks and Hidden Markov Models. Recognizing driving maneuvers is also the goal in [5]. The authors define 27 basic maneuvers between two cars like 'cut in' or 'following', that are compactly modeled using Object-Oriented Bayesian Networks. The work presented in [6] focuses more on prediction: For two cars at an intersection their motion trajectories are predicted in order to foresee possible conflicts.

The approaches of the second category target either highway scenarios or limit the number of considered entities and are therefore not directly applicable to a comprehensive situation assessment in urban scenarios. The approaches of the first category are challenged by the high variability and intricacy of assessing situations. In this paper we propose to tackle these challenges from two sides: Situations are decomposed into parts that can be analyzed individually and are therefore expected to be easier to handle. Additionally, in order to make the recognition of parts computationally feasible, we employ an active measurement process based on computing the mutual information of taking specific measurements. The latter aims at limiting the amount of sensor measurements taken for correctly identifying a part to the minimum necessary. The concept of an active measurement process has already been applied successfully in the field of robotics [7] and in works on probabilistic sensor fusion [8].

In this paper we present our approach along with an evaluation demonstrating both its feasibility as well as its benefits. Section II describes what defines a 'part' in a situation and how the decomposition of situations into such parts is accomplished. In Section III a method for recognizing configurations using Bayesian Networks is presented and the active measurement process is detailed. Section IV gives the results of a thorough evaluation of our approach in simulations and in Section V future research directions for further enhancing the presented methods are discussed.

II. DECOMPOSING A SITUATION INTO ITS PARTS

Among the most striking challenges for situation assessment are the high complexity and variability of situations encountered in urban driving. This complexity is especially given in intersection scenarios, where other road users, traffic lights and traffic rules have to be considered. Even if an

M. Platho is with Department of Neuroinformatics, Technical University of Ilmenau, D-98684 Ilmenau, Germany matthias.platho@tu-ilmenau.de

J. Eggert is with Honda Research Institute Europe GmbH, D-63073 Offenbach am Main, Germany jeggert@honda-ri.de

intersection situation consists only of a moderate number of lanes and vehicles as depicted in Figure 1, assessing the given situation is non-trivial.

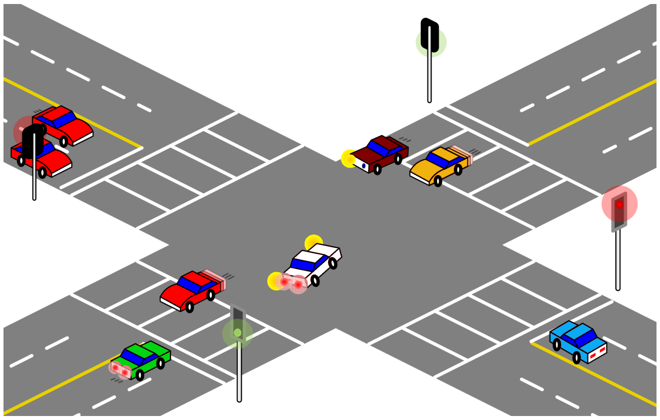

Fig. 1. A typical intersection situation. Although the number of road users and lanes is quite moderate, the assessment of this situation is non-trivial.

Interpreting the whole situation based on learned prototypical situations is hardly feasible, as situations are highly variable in number, constellation and dynamics of the participating entities (cars, pedestrians, traffic lights, etc) and accordingly the state space is very large. At the same time, for a given road user not all of the present entities are equally relevant. E.g. in Figure 1, from the green car's view, the vehicles waiting at the red traffic lights are currently of no direct relevance and will thus not affect its immediate behavior. The entity actually affecting the green car's behavior most is the white car in the center of the intersection. The white car has stopped in order to yield to oncoming traffic and blocks the green car's way, which is thereby forced to slow down.

The preceding description suggests that a situation can be analyzed based on the way the present entities are interrelated. Instead of interpreting or classifying a situation as a whole, the individual vehicles are examined to determine which entities affect their current behavior. This approach decomposes a situation into multiple parts, with each part consisting of a road user and all entities affecting its behavior. We specify such a part in a representation structure that we term *configuration*.

A configuration is defined by its participating entities, the affected entity tagged *reference entity* and its affecting entities (currently we consider only one). Additionally the relations of the entities are specified, which serve as features for indicating the presence of the configuration. Unary relations describe the dynamic states of the entities, e.g. the velocity or acceleration of a car or the state of a traffic light. Binary relations describe relative states between reference and affecting entity like distance or relative velocity between vehicles. If in a given situation the relations of two entities match the relations of a certain configuration, the entity matching the reference entity is said to be in that configuration. The configurations used in this paper were designed by hand; learning them will be part of future work.

An exemplary decomposition of the intersection situation shown in Figure 1 into basic configurations is depicted in Figure 2. The white car is about to turn in the upper left arm of the junction, but has to yield to the approaching orange car and therefore stops. This case can be described by a configuration that is tagged "Stopped by intersection", with the white car being the affected entity (reference entity) and the orange car being the affecting entity. Besides road users also stationary entities like crosswalks and traffic lights can participate in a configuration. In the given situation, the blue car has halted because the traffic light for its lane has turned red. In this "Stopped by red traffic light" configuration the blue car is the reference entity and the traffic light is the affecting entity. The third type of configuration present is "Stopped by leading vehicle": As already described, the green car has to slow down as its way is blocked by the stopped white car. This case also demonstrates that a road user can be both affecting as well as affected entity at the same time (but in different instances of configurations). An entity can even be the reference entity in multiple configurations at once: A car waiting behind the blue vehicle would be both stopped by the latter and the red traffic light.

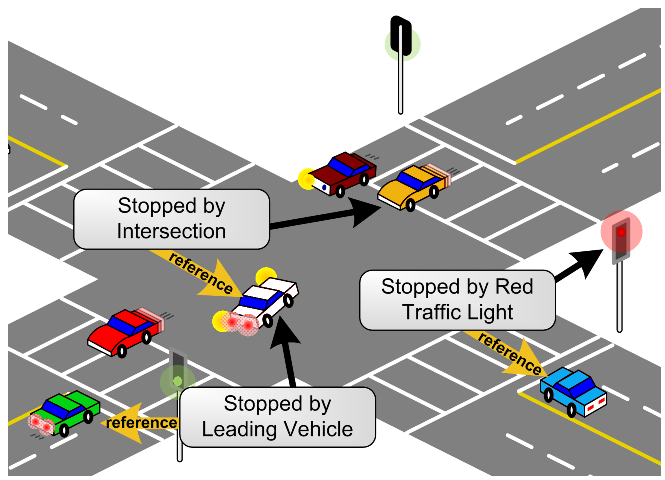

Fig. 2. An exemplary decomposition of the intersection situation into configurations. Each configuration contains an affecting and an affected (reference) entity.

In this exemplary decomposition no configuration was assigned to the two red cars. The reason is that we currently only consider configurations that force the reference entity to slow down or stop. This serves two purposes: First, such configurations are especially safety-relevant, e.g. anticipating the breaking of a vehicle helps to avoid rear-end collisions. Secondly, these behaviors can often be explained by directly or indirectly measurable causes like a blocking obstacle or a certain traffic rule. In this sense the red cars are neither affected by nor affecting any entity and are therefore not part of any of the introduced configurations.

As configurations specify relational aspects they can be naturally represented as graphs. In Figure 3 a simplified sketch of the "Stopped by leading vehicle" configuration is depicted. The root node provides the label of the configuration; its child nodes denote reference entity and affecting

in: Proc. 15th Int. IEEE Conf. on Intelligent Transportation Systems (ITSC 2012), Anchorage, pp. 888-893, IEEE 2012

entity. Relations are represented as child nodes of their corresponding entities. In this example the configuration is present if the reference car is either slowing down (defined in node "Acceleration") or stopped ("Velocity"). It furthermore requires that the leading vehicle is not driving faster than the reference car ("Relative Velocity") and that they are in a certain range ("Distance"). The actual values are not shown in the Figure for simplicity.

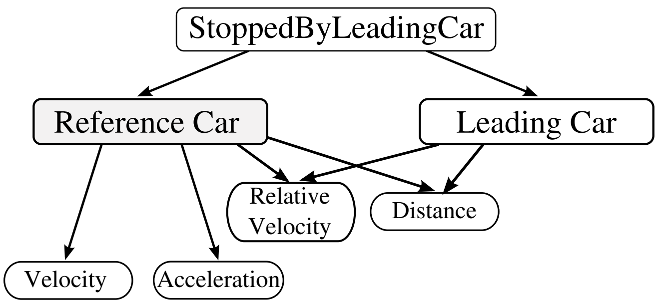

Fig. 3. A schematic representation of the configuration "Stopped by leading vehicle". It contains both unary relations (Velocity, Acceleration) as well as binary relations (Relative Velocity, Distance).

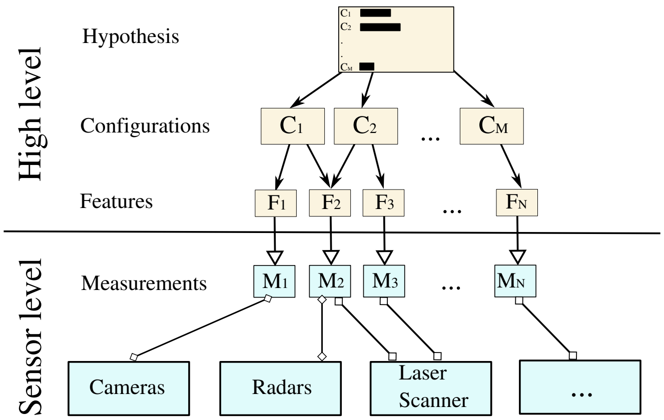

Fig. 4. Schematic representation for the relation between high level and sensor level in a complete system for configuration recognition. The Bayesian Network in the high level relies on measurements taken in the sensor level.

III. ACTIVE MEASURING FOR CONFIGURATION RECOGNITION

For recognizing configurations Bayesian Networks [9] are employed. A Bayesian Network is a probabilistic graphical model for representing random variables along with their conditional dependencies in a graph structure. Nodes denote random variables and directed edges between nodes state conditional dependencies. Bayesian Networks are widely used for probabilistic modeling as they possess various desirable properties. For example, they offer a direct way to incorporate expert knowledge and are able to cope with missing evidence. They are especially suited for our approach since the graphical representation of a configuration can be almost directly mapped to a Bayesian Network and we can use its robustness to missing data.

A schematic example of a Bayesian Network for configuration recognition is shown in the upper half of Figure 4. Each configuration is modeled as a Bayes classifier, consisting of a configuration node C_m and a set of features nodes F_n. Each feature node corresponds to a single relation of the configuration and the configuration node provides the belief in the presence of the configuration. A single, superordinated hypothesis node combines the beliefs of all configuration nodes in one place. The configuration having the highest probability in the hypothesis node is returned as result for the regarded road user.

The lower part of Figure 4 depicts the sensor level as it would be required for a complete recognition system. It is capable of performing different measurements for perceiving the environment, each of which requires reading one or multiple sensors and performing a suitable computation. Taking a measurement usually generates some kind of costs, e.g. by blocking an exclusive resource like a pan-and-zoom camera, occupying a data bus or requiring a demanding computation. Even if no costs are directly generated, at least a certain amount of time has to be spent to wait for the result.

The high level comprising the Bayesian Network and the sensor level are only connected in one way: For obtaining the value of a certain feature in the Bayesian Network the corresponding measurement has to be taken in the sensor level. For each individual feature a dedicated measurement is required. Apart from that, high level and sensor level are independent.

When comparing the generally marginal computational costs generated by inference performed in the Bayesian Network with the costs from measuring a feature, the latter requires usually significantly more resources. This means that trading computations in the high level against taking measurements in the sensor level would be advantageous. This is indeed possible when employing an active measurement process.

The general idea of such a process is to actively trigger measurements based on the expected gain for a given task. The goal is to save measurements by limiting the measurements taken to the necessary ones for a specified level of performance. When using probabilistic models like Bayesian Networks the selection of a suitable measurement to take can be based on its expected mutual information [10]. Employing this method to situation assessment is an additional step for dealing with the complexity of situations encountered in urban driving that complements the decomposition method described in the previous chapter.

The active measurement process proposed here measures features sequentially, one after the other, until the probability of a hypothesis for a certain configuration (including 'No configuration') surpasses a predetermined threshold τ. The process pursues two goals: First, to incrementally gather evidence in a way that reduces the set of probable hypotheses quickly to one. Secondly, to terminate the costly measurement process as soon as it becomes improbable that further measurements will improve the current hypothesis. Its working principle can be summarized in four steps:

in: Proc. 15th Int. IEEE Conf. on Intelligent Transportation Systems (ITSC 2012), Anchorage, pp. 888-893, IEEE 2012

1) *Measurement Selection*: Evaluate information gain of yet unobserved features.
2) *Observation*: Measure the feature which provides the highest expected information gain, i.e. which is expected to decrease the uncertainty on the current configuration hypotheses most.
3) *Inference*: If the belief in the most probable configuration is lower than τ and not all features have been measured already, continue with 1, else continue with 4.
4) *Result*: Return the most probable configuration.

The information gain is determined based on computing the mutual information I between hypotheses H and features $F_i, i \in 1, ..., n$ with n being the number of features. The mutual information $I(H; F_i)$ is defined as:

$$I(H; F_i) = \sum_h \sum_f p(h, f) log \frac{p(h, f)}{p(h)p(f)} \tag{1}$$

where $p(h)$ stands for the probability of hypothesis h and $p(f)$ stands for the probability of F_i having value f. The mutual information provides the expected reduction in entropy of H given an observation of F_i. The higher the mutual information is between two random variables, the more knowing one tells about the other. In the case of continuous features either a discretization can be applied, e.g. using the method presented in [11], or the summation can be matched with a definite double integral while the probability distribution functions are replaced by probability *density* functions.

The probabilities needed for these computations are provided by the Bayesian Network, which in turn obtains its prior and conditional probabilities by parameter estimation on a set of training cases.

An important aspect that can not be directly seen from Equation 1 is the fact that the mutual information between two variables changes if evidence on dependent variables is entered. So in order to always measure the feature with the highest information gain the mutual information has to be recomputed after each observation, as in our Bayesian Network the features are not independent from each other. This means that the order of measurements has to be determined online and can not be computed beforehand (offline).

IV. EVALUATION

As the feasibility of decomposing situations into configurations and reliably recognizing them is already demonstrated in related work [12], this evaluation focuses on quantifying the benefits of an active measurement process. In order to test the process we investigate how good it performs on its two strongly related subtasks: Incrementally increasing the confidence (belief) of the correct hypothesis with as few measurements as possible and achieving reliable recognition rates with also as few measurements as possible. The former task is important because the confidence is used in the measurement process as stopping criterion.

For the evaluation a simulated intersection scenario, consisting of a 4-lane major road crossed by a 2-lane minor road, was set up. The intersection is regulated by traffic lights. Cars approach from all directions and, depending on their goal and lane, go straight or turn left or right. The scenario is simulated in a microscopic simulation framework which was developed in the course of the research on situation decomposition; a screenshot of the simulation is shown in Figure 5. The framework employs standard models for e.g. car-following [13] and was created in order to gain direct access to the driver agent level, which is hard to obtain in most commercial traffic simulators. This access is crucial for our approach, as for determining whether a configuration was correctly recognized the information which entities affected a driver's behavior is needed.

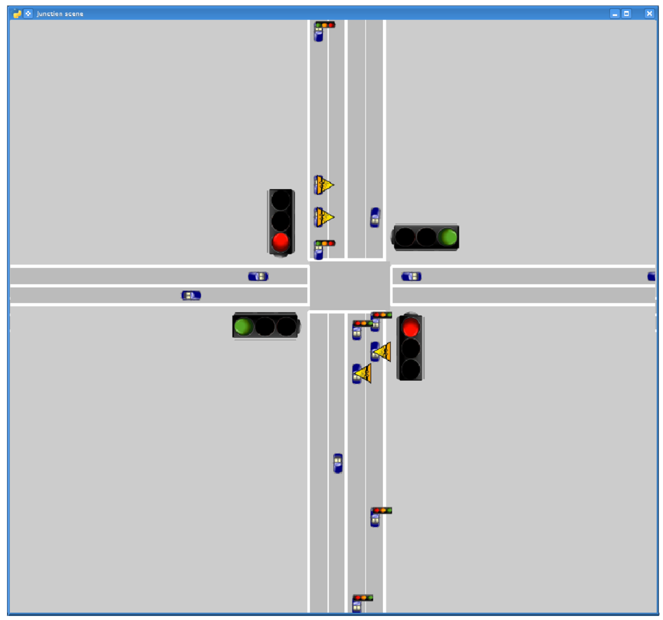

Fig. 5. Screenshot from the simulated intersection. The symbol above a car denotes the configuration it is currently in. Cars without symbols are accelerating or driving at constant speed.

The data for the evaluation was generated by running the simulator for about twenty minutes on the intersection scenario. Every 100 milliseconds the dynamic state of each present entity was logged, for cars additionally their currently affecting entities were recorded. Based on the recorded information, for each car at each recorded time instance a case was generated, consisting of the measurements of the configurations' features and the current configuration. A total of 142030 cases were generated. For the reported results tenfold cross validation was used.

The Bayesian Network used for recognizing configurations is depicted in Figure 6. Like in the schematic Bayesian Network described in Section II the configuration nodes (colored green) provide the belief of being in the corresponding configuration. Each configuration node possesses two unique features while two features are shared among all configurations: *Velocity* and *Acceleration* of the reference entity. The reason for using these features for all configurations is that

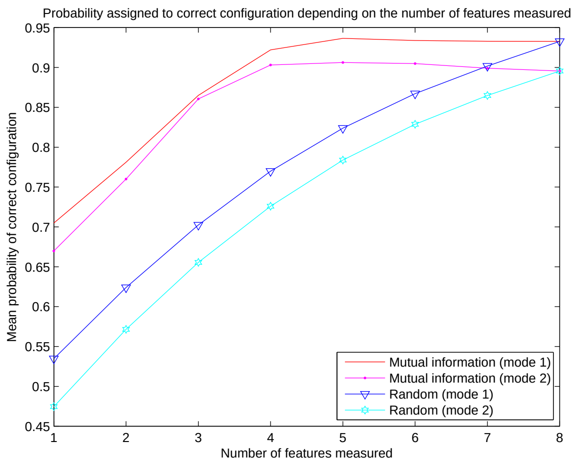

Fig. 7. The plot depicts the different progress of the belief in the correct configuration for both variants of measurement selection (based on mutual information versus random). An active measurement process based on mutual information achieves a high belief already after about half of the possible measurements.

according to the definition given in Section II the reference entity in a configuration has to be either braking or stopped, otherwise it is in no configuration. The topmost node serves as decision node as it combines the beliefs for the individual configurations and its most probable hypothesis (including 'No Configuration') is taken as result.

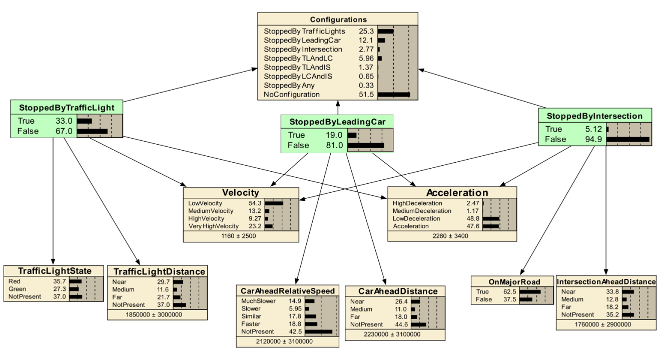

Fig. 6. The Bayesian Network that was used for the second mode of the evaluation. The configuration nodes are colored green. In the network for the first mode the "Configurations"-node has only 4 states (i.e. the basic states 'StoppedByTrafficLight', 'StoppedByLeadingCar', 'StoppedByIntersection' and 'NoConfiguration'), but no states for combined configurations.

There are cases in which a car is in multiple configurations at once ('combined configurations'). Generally, one is interested only in the determining configuration, which is the one causing the highest deceleration of the car. For example, if a car approaches a distant red traffic light at medium speed it needs to decelerate only slightly. If at the same time a vehicle directly ahead performs an emergency brake the car is forced to brake hard. In this case the determining configuration would be 'StoppedByLeadingCar'. Nevertheless the information about the 'StoppedByRedTrafficLight'-configuration might still be useful later on so recognizing it would be of additional benefit. This is why all of our following evaluations are performed in two modes. In mode one the goal is to recognize only the determining configuration and in mode two all current configurations, including combinations of the basic three, have to be recognized.

In order to quantify the gain from measuring features incrementally based on their mutual information, an additional method was employed for comparison, where the sequential order of measurements is determined randomly.

Belief in correct configuration

In this part of the evaluation the benefits of the proposed active measurement process over a measurement process with a random sequential feature selection are investigated. The expectation is that the active measurement process achieves a high belief in the correct configuration at an early stage with much fewer measurements. The results are shown in Figure 7. The x-axis refers to the number of measurements taken, the y-axis refers to the belief in the correct configuration.

For both modes the measurement selection based on mutual information leads to significantly higher beliefs in the correct configuration than the random selection does - up to 20% points. When employing the random selection process the belief tops 80 percent not until 5 (mode 1) or 6 (mode 2) measurements have been taken. Opposed to that, the process guided by information gain tops it after only 3 (3) measurements. The proposed measurement process thereby proves to obtain a higher confidence in the correct configuration than the random process while taking significantly less measurements.

Recognition rates

The next question to investigate is whether the incremental, active measurement process can reliably recognize configurations while measuring only a subset of the available features. Figure 8 depicts the results: The x-axis refers again to the number of measurements taken and the y-axis gives the percentage of correctly recognized configurations.

Using mutual information as selection criterion results in a steep increase of the recognition rate as compared to the gradual increase of random selection. After taking only three measurements the recognition rate for both modes surpasses 90% when feature selection is based on expected information gain, while the random process obtains less than 80%. Employing an incremental measuring process thereby saves measuring more than half of the features without impairing the recognition rate more than 1.5% points.

V. DISCUSSION AND FUTURE WORK

The results presented in this paper demonstrate the gain from employing an active measurement process, as it vastly reduces the effort for recognizing configurations. Configurations in turn provide a way to decompose situations into their constituent parts. Put together, these two methods provide a viable approach for tackling the high complexity and variability of situations encountered in inner-city scenarios by systematically reducing the computational effort

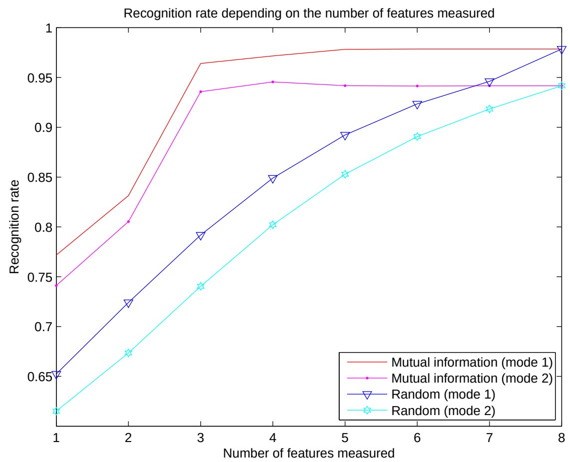

Fig. 8. This plot compares the recognition rates achieved after measuring one to eight features when selecting the order of measurements either based on mutual information or randomly. It shows that the active measurement process (mutual information) achieves reliable recognition rates when measuring only three features. Interestingly, the recognition rate in mode 2 (mutual information) drops after taking 5 measurements. A possible explanation is that some features are unreliable, but this has to be further investigated.

needed for situation assessment. The active measurement process facilitates an incremental recognition of a road user's configurations, while recognizing configurations allows for incrementally building up an understanding of the current situation.

In future work, the active measurement process will be further extended to circumvent testing configurations on all pairs of entities. Instead, based on feature priors only reasonable pairs should be examined.

The evaluation was performed in a simulation framework, where erroneous sensor measurements are not yet considered. In order to achieve a sufficient robustness to sensor noise as required by a real-world system, multiple approaches will be investigated, e.g. modeling sensor readings probabilistically or enabling the detection of conflicting measurements, which can also be accomplished in the Bayesian Network framework currently used.

At this stage, the approach answers solely the question *'Who is affected by what?'*. Nevertheless, this provides a basis for situation assessment, that will be build on in future work. One major research target will be to predict the progress of a situation based on recognized configurations. A potential approach could consist of two parts: A method to anticipate when a road user will enter or leave a configuration and a method for 'linking' existing configurations in order to incorporate indirect interrelations. *Linking* means to consider transitivity in configurations: If vehicle A affects vehicle B, and vehicle B affects vehicle C, then C is indirectly related to A. We expect that these methods will provide a more comprehensive understanding of a given situation and accomplish a major step towards building a driver assistance system for inner-city driving.

References

[1] S. Vacek, T. Gindele, J. M. Zöllner, and R. Dillmann, "Situation classification for cognitive automobiles using case-based reasoning," in *Proc. IEEE Intelligent Vehicles Symp*, 2007, pp. 704–709.

[2] M. Hülsen, J. M. Zöllner, and C. Weiss, "Traffic intersection situation description ontology for advanced driver assistance," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, 2011, pp. 993–999.

[3] R. Schubert, K. Schulze, and G. Wanielik, "Situation assessment for automatic lane-change maneuvers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 607–616, 2010.

[4] D. Meyer-Delius, C. Plagemann, and W. Burgard, "Probabilistic situation recognition for vehicular traffic scenarios," in *Proc. IEEE Int. Conf. Robotics and Automation ICRA '09*, 2009, pp. 459–464.

[5] D. Kasper, G. Weidl, T. Dang, G. Breuel, A. Tamke, and W. Rosenstiel, "Object-oriented bayesian networks for detection of lane change maneuvers," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, 2011, pp. 673–678.

[6] E. Käfer, C. Hermes, C. Wöhler, H. Ritter, and F. Kummert, "Recognition of situation classes at road intersections," in *Proc. IEEE Int Robotics and Automation (ICRA) Conference*, 2010, pp. 3960–3965.

[7] R. D. Rimey, "Control of selective perception using bayes nets and decision theory," Ph.D. dissertation, Rochester, NY, USA, 1994.

[8] Y. Zhang and Q. Ji, "Active and dynamic information fusion for multisensor systems with dynamic bayesian networks," *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, pp. 467–472, 2006.

[9] J. Pearl, "Bayesian networks: A model of self-activated memory for evidential reasoning," in *Proceedings of the 7th Conference of the Cognitive Science Society, Irvine*, Aug. 1985, pp. 329–334.

[10] C. Shannon and W. Weaver, *The Mathematical Theory of Information*. University of Illinois Press, 1949.

[11] U. M. Fayyad and K. B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning." in *IJCAI*, 1993, pp. 1022–1029.

[12] M. Platho, H.-M. Groß, and J. Eggert, "Traffic situation assessment by recognizing interrelated road users," in *Proc. IEEE Intelligent Transportation Systems (ITSC)*, 2012.

[13] W. Helly, "Simulation of bottlenecks in single-lane traffic flow," *Theory of Traffic Flow*, pp. 207–238, 1961.