2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

A General Reliability-Aware Fusion Concept Using DST and Supervised Learning with Its Applications in Multi-Source Road Estimation

Tran Tuan Nguyen[1], Jens Spehr[1], Dominik Vock[1], Marcus Baum[2], Sebastian Zug[3], and Rudolf Kruse[3]

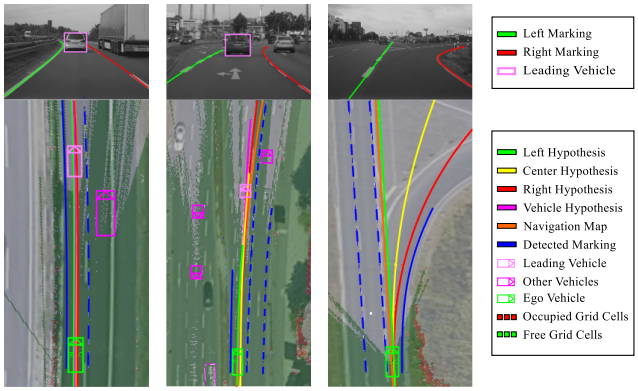
***Abstract*— Concerning the variety of challenging situations of automated driving, a simple average fusion of all available information sources is not appropriate to obtain satisfying results. Hence, this paper presents a novel framework for the road estimation task by incorporating reliability into the multi-source fusion. First, we specify the common JDL fusion model for this task and extend it at multiple levels. Secondly, we integrate an offline-trained knowledge base for the reliability assessment represented by Bayesian Network or Random Forests. Thirdly, we propose a reliability-aware fusion of various sources at the decision level by applying Dempster-Shafer theory. As a result, our system can solve conflict situations among the sources more satisfyingly. Compared to the average fusion, experiments on real world data verify that our concept can increase the overall performance of automated driving up to 8 percentage points.**

I. Introduction

A. Motivation

The key requirement for automated driving is an environment perception that can cover all relevant situations. Due to the large bandwidth of the operational scenarios (e.g., traffic, weather, road conditions), various sensor types and/or feature detectors have to be combined to cope with their individual limits and disadvantages. In this process, an optimal fusion strategy should weight the outputs of the different sources appropriately to obtain the best results. For applications with limited variability of environmental conditions, acceptable results can be achieved by using a specific set of static weights regarding a failure model for each scenario. For outdoor and especially automotive applications, constant failure models for the sensors and detectors cannot be assumed. Hence, a scenario-sensitive weighting method which can determine the quality of the given sources at runtime has to be designed.

For this purpose, Rogova et al. describe explicit and implicit approaches [1]. Based on decision systems and hand-designed rules, explicit methods such as in [2] provide fixed weights for the fusion. By that, the weights represent the vectorized quality information or failure categories and frequencies of the sources for each scenario. Independent of the sensor setup, these methods provide a global and objective policy for weighting the sources [3]. However, they require a significant conceptual and computational effort to generate comparable quality information. In contrast, implicit methods apply machine learning approaches to obtain the reliability knowledge [1]. At design-time, the knowledge is derived from training data by evaluating reference values. The advantage of these methods is that the learning process can be highly automated. Their drawback, however, lies in the situation-specific definition of the knowledge base. For example, raw camera images are directly mapped to steering commands by using end-to-end learning systems [4]. Such methods are computationally expensive and require a huge amount of training data. Besides the estimation of reliabilities, their integration into the fusion process has not been widely investigated [1], [5]. Therefore, our work presents a reliability estimation method by combining the advantages of both these approaches. Moreover, we introduce a novel way of integrating reliabilities into the fusion. As a proof of concept, we apply our introduced concept to the problem of multi-source road estimation.



(a) Highways. (b) Urban road. (c) On-ramp.

Fig. 1: Different situations with detection results and the model-based ego-lane estimations (*Google Maps*™).

B. Application Scenario

For road estimation task, the dominating sensor type is camera [6]. The performance of cameras, however, is highly influenced by weather, lighting conditions and the quality of the lane markings. Hence, we integrate further orthogonal information sources, e.g., leading vehicles and digital maps, to compensate the drawbacks of each individual source. Using a static sensor setup, our road estimation framework can generate various ego-lane estimations such as

- driven path of leading vehicle: *Vehicle Hypothesis (VH)*
- solely left lane marking: *Left Hypothesis (LH)*
- solely right lane marking: *Right Hypothesis (RH)*
- both lane markings: *Center Hypothesis (CH)*

[1]Volkswagen Group, Berliner Ring 2, 38440 Wolfsburg, Germany [firstname.surname]@volkswagen.de
[2]University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany [firstname.surname]@cs.uni-goettingen.de
[3]Otto-von-Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany [firstname.surname]@ovgu.de

Fig. 1a shows that the hypotheses are usually reliable on highways so that any choice would be correct. In urban scenarios, however, there are more situations where the hypotheses disagree (Fig. 1b), e.g., poorly visible lane markings, leading vehicle starts to change the lane, etc. In such cases, the appropriate set of hypotheses has to be identified to keep the vehicle in the lane. In critical situations with only one correct hypothesis (Fig. 1c), a correct reliability estimation of the hypotheses is essential.

The contributions of our *reliability-aware fusion* concept are presented by applying it to the multi-source fusion problem in the field of road estimation. A detailed survey about these topics is presented in Section II. In Section III, we adapt the JDL fusion model to road estimation and specify the components for each level. In Section IV, we introduce a novel reliability estimation process, which is performed by learning from the training data. By that, different features are used, e.g., sensor-related, contextual and consensus information. Hence, our system becomes scalable to new scenarios. Compared to our previous works [7], [8], this paper integrates further features, such as free space information or a rough road curvature from a navigation map. To combine the hypotheses, we incorporate the predicted reliabilities and the classifiers' performance to the fusion using the Dempster-Shafer theory. In this way, our fusion can solve the conflict situations among the sources more satisfyingly (Section V). Furthermore, our data fusion concept can be generalized to any high-level competitive fusion, e.g. for in- and outdoor robot navigation.

II. Related Work

Before our concept is presented in the next chapter, this chapter provides a short review about the related works in the field of road estimation and incorporating reliabilities.

A. Road Estimation Using Multiple Sources

For robust road estimation, multiple works [6] fuse the information from different sources. These approaches can be divided into two categories: *low-level* and *high-level* fusion.

1) Low-level: At low levels, the output of several sensors is directly combined. For example, lidar scan points are projected into the camera images to reduce false positives [9] or used to limit the region of interest for later image processing [10]. Moreover, after the image registration, geometric and color features from lidar and camera are fused to classify the drivable areas [11]. This combination of independent sources towards resolving information incompleteness is also known as *complementary fusion* [12]. Its complexity, however, grows rapidly with the number of sources since the data association among the sensors depends on the quality of the coordinate transformation and cross calibration. Besides, such vision-based methods are not robust against lighting and weather conditions and do not involve redundant information from other sources such as leading vehicles.

2) High-level: Here, the output of the sources often has a model-based representation, e.g., clothoid model [13], spline [14], grid-based model [15], etc. Afterwards, the independently generated models from various sources are combined in a *competitive fusion* [12] to minimize the errors between the estimation and the real road course. For instance, the coefficients of the clothoids are averagely fused when combining information of lane markings, GPS or other vehicles [16]. Since this focuses on minimizing spatial errors, they cannot handle conflict situations satisfactorily due to the assumption that all sources are equally reliable.

B. Incorporating Reliability in Fusion

Because the success of the fusion depends on the quality of the data uncertainty model [1], this section introduces related work of defining, estimating and combining *reliability*.

1) Introduction of reliability: In many works, reliability is related to the terms *confidence* [17] and *trust* [18]. It denotes the state of uncertainty between the detection and the real object's state, which is also known as the first-order uncertainty, e.g., spatial imprecision [19],likelihood functions [20], detection accuracy [21], etc. However, this kind of uncertainty is also an estimation, therefore *a second-order uncertainty* is needed to indicate the stability of the first-order uncertainty [17]. According to this definition in [17], *reliability* of a hypothesis h in this work is denoted as the *reliability coefficient* R_h.

2) Deriving reliability: The simplest approach to obtain R_h of a hypothesis h is by using expert and domain knowledge [2]. Unfortunately, it is often subjective and hard to define an appropriate mapping function for R_h. Alternatively, the consensus among all available hypotheses can be employed. For example, a neural network is trained to evaluate the correctness of the navigation map by comparison with sensor-based detections [22]. By that, their network can only detect the existence of errors, but is not able to identify the incorrect hypotheses. Similarly, a related approach is applied in [18] to assign R_h to each h: the more hypotheses agree with h, the greater R_h is. Indeed, the reliable hypothesis can be outvoted by the unreliable ones. In [23], the reliability is determined by a discounting rule of belief theory. They estimate the discounting by minimizing the mean squared error of a distance metric between the sensor data and the reference. The smaller R_h, the larger the discounting. This can lead to a high computational cost in case of a large number of sensors. Alternatively, the authors in [24] assign R_h to each object detector regarding its discrepancy to the reference in the offline training phase using a machine learning approach. However, they disregard the contextual information which are essential to the general fusion as proposed by [25].

3) Usage of reliability: Approaches towards integrating the estimated R_h into the fusion can be divided into three groups. The most commonly used approach is to use R_h as weight for a weight-based fusion as in [23], [24]. Another strategy is to exclude all unreliable sources from the average fusion [26]. Besides, those methods can be combined to a weight-based fusion excluding the unreliable sources.

To solve the discussed problems, we apply supervised learning approaches to obtain reliabilities. Based on the reliability information, the high-level competitive fusion can solve conflict situation and achieve better performance.

III. Overall Concept

The previous sections clarifies the relevance of fusing multiple sources for road estimation. By that, a proper incorporation of reliabilities can leverage the fusion's performance.

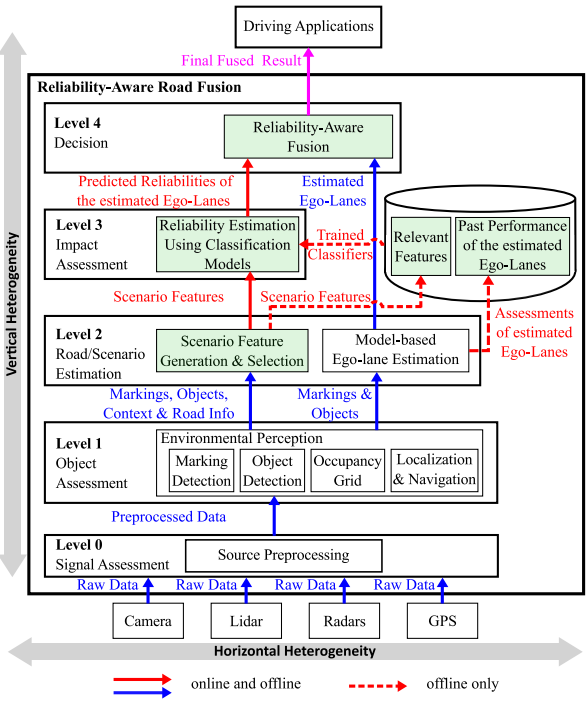


Fig. 2: Reliability estimation and reliability-aware fusion as an additional supervision system within the road estimation task (Blue: Data for road detection; Red: Reliability information).

Regarding the fusion model from the US Joint Directors of Laboratories (JDL) [27], Fig. 2 illustrates our concept of a reliability-aware road estimation consisting of multiple levels. In Level 0, the raw sensor data is preprocessed on the basis of physical signal level. In Level 1, multiple detection modules iteratively utilize the preprocessed data to estimate and predict the states of different object types. This includes tasks such as object detection, tracking, association, etc. The low-level fusion, e.g., object association of different sensors, is taking place here. These two levels form the *Perception Layer* for which many works have been published in recent years [6], [28]. In our work, the used sensors are delivered with their internal processing modules and provide different results such as lane markings, dynamic objects, etc. They apply state-of-the-art methods [6] and are not discussed further.

In Level 2, a model-based lane estimation is performed by using a modified version of the compositional hierarchical model from [21]. This process utilizes the detections from Level 1 by exploiting their spatial and temporal relationships. This results in four different ego-lane estimations ($H = \{LH, RH, CH, VH\}$), which are generated independently without cross checks among each others. A problem arises in conflict situations between the ego-lane hypotheses, i.e., pointing in different directions. In this case, the decision layer needs additional information to make correct choices.

To provide a scalable and adaptable solution for this problem, we introduce at Level 3 a *reliability estimation* module, which is based on classifiers. Following the definition of reliability as the *second-order uncertainty*, this module is related to threat assessment of the JDL model and represents the consequences of choosing each hypothesis. For this purpose, we add a new component at Level 2 which extracts relevant information from the sensor detections. The extracted results are called "*scenario features*" and represent different types of information, e.g., road type (highways, urban areas), consensus among the hypotheses, occupancy of a lane, etc. This extraction component at Level 2 does not make any interpretation about the impact of choosing each hypothesis because this task is taking place in the reliability estimation. In the offline phase, these features are combined with the past performance of the hypotheses to train several classifiers. After, these trained models are used to predict the reliabilities online. In Level 4, these reliabilities support the *reliability-aware fusion* so that unreliable sources will be rejected and no longer affect the fusion result. Compared to many works such as [9], [10], this fusion is taking place at a high level and not at the sensor level. Based on the predicted reliabilities, our fusion concept acts as an additional supervision system to select the best ego-lane estimations for self-driving functions. The following sections will discuss our concept in detail.

IV. Data-Driven Reliability Estimation

This section discusses our concept of reliability estimation using training data. Afterwards, the structure of these data and the applied classifiers are presented in detail.

A. Concept

As discussed, there are two main approaches to estimate reliabilities. The explicit approaches require a huge effort for the design and maintenance of the system, e.g., sensor update. In contrast, the implicit methods such as end-to-end learning approach are not useful for our problem of choosing the most reliable estimations because the complexity of the learning problem increases with the number of sources. Moreover, it is not possible to access the raw data of all sensors.

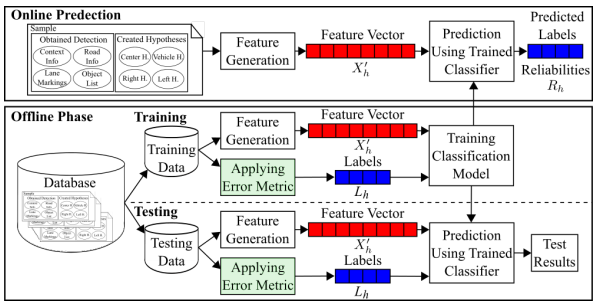


Fig. 3: Concept for learning and predicting reliability. Labels are generated using the error metric presented in Section IV-C.

To take advantages of both methods, we model the reliability estimation as a classification problem with the extracted *scenario features* as input. By that, the perception layer serves simply as a black box and we are not interested in its internal processing. Fig. 3 illustrates our concept with two running modes. In offline mode, a database is built with the feature vector X_h as observation for a hypothesis h and

the corresponding past performance of h as its label l_h. By that, the performance of h is obtained by applying an error metric, which measures the difference of h to the reference (Section IV-C). By assessing the obtained difference, the label l_h of each h is then assigned to one of two classes: *Reliable* or *Unreliable*. In the online mode, such as on-road or testing applications, the trained classifiers predict the class membership probability for each ego-lane h. This probability is then used as the reliability coefficient R_h. In the following sections, each step will be described in greater detail.

B. Generation and Selection of Scenario Features

This section describes the feature vector extracted from the perception layer's output. To predict the reliability of each hypothesis h, the complete set of all features consists of three different groups with $X_h = [s_h, \tau_h, \gamma_h]$ as in Fig. 4.

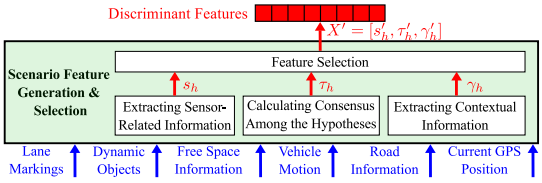


Fig. 4: Generation and selection of scenario features.

The first group s_h denotes the sensor-related information of h, e.g., the clothoid parameters of the left and right lane markings (*LM* and *RM*) and the driven trajectory of the leading vehicle (*VEH*). In our work, a clothoid is approximately represented by $y(x) = y_0 + \phi x + \frac{1}{2}c_0x^2 + \frac{1}{6}c_1x^3$ with y_0 (initial lateral offset), ϕ (angle to the x-axis), c_0 (curvature), c_1 (curvature change). Besides, the internal processing modules of the sensors append an existence value ξ to each detection. In contrast to the state uncertainty (e.g., position, size), ξ represents the uncertainty whether the detected object really exists [29]. In our framework, ξ comes from different uncertainty theories so that ξ is not directly comparable. Besides, we use the occupancy grid to check if h lies in a free area with no obstacles. This results in the feature $free_h$ that expresses the amount of free cells along the clothoid representing h. Moreover, Ib shows further information of s_{VH} such as position, velocity, etc.

The second group τ_h represents the consensus among all hypotheses (Tab. Ia). This is measured by comparing the deviation of the clothoid parameters representing *LM*, *RM*, *VEH*, with their average $(\bar{l}, \bar{\phi}, \bar{c_0}, \bar{c_1})$. These features reflect whether the hypotheses agree among themselves or there are conflicts between them. Hence, τ_h considers the dynamic aspects, whereas s_h focuses on the static aspects.

The last group γ_h describes the contextual features, which in turn are divided into the internal $\gamma_{h,int}$ and the external environment parameters $\gamma_{h,ext}$. Thereby, $\gamma_{h,int}$ describes the motion of the ego-vehicle (Tab. Ib). Besides, $\gamma_{h,ext}$ contains the information extracted from a navigation map such as the *Road Class* (e.g., highway, rural, urban, connection) or the *Lane Type* (e.g., normal, ramp, roundabout, split, merge, etc.). Another feature of $\gamma_{h,ext}$ is the estimated lane width $Lane_w$.

Overall, X_h consists of 54 scenario features. Certainly, some of them are redundant or irrelevant for the reliability

Feature	Description	Feature	Description
$CS_h\ c0$	Deviation to $\bar{c_0}$	$CS_h\ c1$	Deviation to $\bar{c_1}$
$CS_h\ l$	Deviation to $\bar{l}$	$CS_h\ \phi$	Deviation to $\bar{\phi}$

(a) Consensus features τ_h ($h \in \{LM, RM, VEH\}$).

Feature	Description	Feature	Description
$o\ x$	x-position	$o\ v_x$	Longitudinal velocity
$o\ y$	y-position	$o\ v_y$	Lateral velocity
$o\ v_\phi$	Yaw rate	$o\ v_{LM}$	Velocity to LM
$o\ turn$	left, right, straight	$o\ v_{RM}$	Velocity to RM

(b) Motion parameter of o with $o \in \{EGO, VEH\}$.

TABLE I: Examples of scenario features.

estimation of a specific h. Concerning the *curse of dimensionality* [30], we apply *feature selection* to identify the discriminant features for each h. Otherwise, the classification will have poor performance due to the high-dimensional X_h. In general, there are three approaches towards feature selection: *filter*, *wrapper*, and *embedded* methods [31]. Regarding the moderate performance of *filter* methods and the expensive computations of *wrapper* methods, we apply *Random Forests*, an *embedded approach*, to reduce the feature dimension. The resulted feature sets X_h' are shown in Section VI-A.

C. Automatic Labeling Using Error Metric in Offline Phase

After the previous section explains the composition of the feature vector, this section describes how to get the labels.

To make our trained classifiers scalable to a big amount of real data and consequently easily adaptable to new situations, an efficient labeling with less human effort is required. For this purpose, we apply an error metric to assess the reliability of the ego-lane estimations instead of manually labeling each single one. Regarding the detailed survey of state-of-the-art error metrics in [32], we decide for the metric illustrated in Fig. 5. This metric measures the angle difference $\Delta\alpha$ of an ego-lane estimation h to the reference. Thereby, the reference

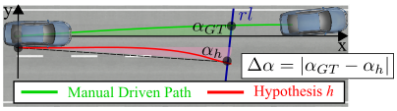


Fig. 5: Reliability as $\Delta\alpha$ between estimated own lane hypothesis and manually driven trajectory [8].

is represented by the manual driven path so that only low-cost velocity and yaw rate sensors are needed instead of the expensive combination of *Differential Global Positioning Systems* and detailed maps. Furthermore, we use the angle difference and not the distance metric because the estimations such as the leading vehicle's trajectory or the estimated lane based on markings will never completely overlap with the reference. In contrast, as long as we drive a nearly optimal path there will be a very small angle difference between the estimations and the reference.

Finally, we use a threshold of $\epsilon_\alpha = 2°$ as proposed in [32] to label the reliability of each estimated ego-lane h as

$$l_h = \begin{cases} 1 \text{ (Reliable)} & \Delta\alpha \leq \epsilon_\alpha \\ 0 \text{ (Unreliable)} & \Delta\alpha > \epsilon_\alpha \end{cases}$$

D. Classifiers for Estimating Reliability

As described in Section IV-A, we model the reliability estimation as supervised learning problems. Hence, this section describes how to obtain the reliabilities using different classifiers trained with the explained observations and labels.

1) Bayesian Network (BN): We apply a *BN* for the reliability estimation because of its advantages such as the mathematical foundation and the comprehensibility due to the graphical representation [30]. Especially, its ability to combine different types of knowledge (e.g., training data, expert knowledge) is well suited for our problem. Using the highly discriminate features $X'_h = [s'_h, \tau'_h, \gamma'_h]$ of each h, we model each $x'_h \in X'_h$ as a discrete node in the *BN* (Fig. 6). Regarding our observations, that the sensor-related features

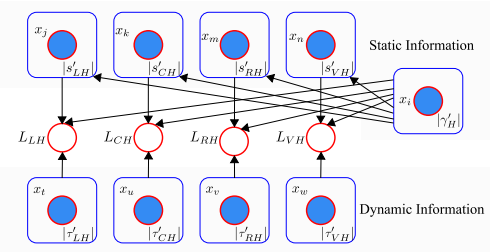


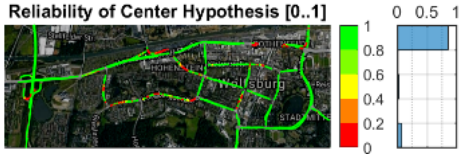
Fig. 6: Modeled Bayesian Network for learning R_h.

depend strongly on the contextual features such as the road class, we model γ'_H as parent nodes of s'_h. For example, the camera can detect lane markings in greater distance on highways than on urban roads. In contrast, the consensus features τ'_h do not have any parent nodes so that the dynamic behavior among the hypotheses is not related to the static information. Finally, the reliability nodes L_h are modeled as independent of each other to simplify the problem. When all processing modules can provide their detection results, all nodes s'_h, τ'_h and γ'_H are observable. Otherwise, when one module is "missing" due to software or hardware failure, the corresponding features are not set and act as hidden variables. After the training, R_h of each h results from the inference through the *BN* with $R_h = P(L_h = Reliable | s'_h, \tau'_h, \gamma'_H)$.

2) Mapping UTM to Reliabilities (MP): Instead of using the extracted features X'_h as training data, this method utilizes the *Universal Transverse Mercator (UTM)* coordinates. By that, the coordinates of the traveled roads are grouped into clusters regarding the euclidean distances. Each cluster C contains $2*|H|$ counters, where $c_{h,Rel}$ indicates the number of reliable observations of h and $c_{h,\overline{Rel}}$ the number of unreliable observations collected from coordinates within the radius r_C. Hence, the reliability R_h for a given position P is calculated by $R_h = \frac{c_{h,Rel}}{c_{h,Rel}+c_{h,\overline{Rel}}}$ regarding the nearest cluster C to P.

As an example, Fig. 7 illustrates the resulting reliabilities of CH and LH from a set of real recordings. It can be seen that LH is more often reliable than CH because the right lane markings on some roads in this area are missing or poorly visible. Besides, we are aware that *MP* and related approaches [26] using real world coordinates cannot be generalized to handle queries from unknown areas.

3) Other Models: For the sake of completeness, further classifiers are applied and later compared with *BN* and *MP*.



(a) Reliability of CH using on Google Maps™.

(b) Reliability of LH using on Google Maps™.

Fig. 7: Reliabilities of CH and LH from a data set ($r_C = 2m$).

a) Random Forests (RF): As one of the mostly used classifiers, *RF* consists of N classification trees [33]. Each of them is constructed with a randomly selected subset of observations and features. One *RF* is trained for predicting each R_h. By that, R_h reflects the amount of trees N_R voting for a reliable h as $R_h = \frac{N_R}{N}$.

b) k-Nearest Neighbors (kNN): Using kNN, the reliability R_h is determined as $R_h = \frac{k_R}{k}$ with k_R represents the neighbors voting for the class *Reliable* among the nearest k neighbors. This differs from *MP* by using the scenario features instead of the UTM coordinates.

c) Support Vector Machine (SVM): *SVM* classifies by finding an optimal hyperplane which maximizes the margin between the classes [34]. We apply the sigmoid function $R_h = 1/(1 + e^{-d})$ to map the distance d between the given feature vector and its next margin to R_h.

V. Reliability-Aware Fusion

A. Concept

Taking place at *Decision Layer* (Level 4), the core of our reliability-aware fusion is the fusion strategy using the estimated reliabilities and *Dempster-Shafer theory (DST)* [35] to solve conflict situations (Fig. 8). To make a strong case

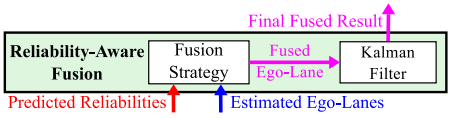


Fig. 8: Reliability-aware fusion consisting of two steps.

of our work, we apply additional strategies and show that these can be outperformed by our strategy. Afterwards, the fused outcome is tracked with a Kalman Filter before the final robust result can be passed to the driving functions.

B. Fusion Based on Dempster-Shafer Theory and Reliabilities

This section utilizes *DST* to combine the estimated ego-lanes using the predicted reliabilities. By that, we integrate the classification performance as the third-order of uncertainty.

First, we cluster the available hypotheses into various groups $G_{1..N}$ so that the angular differences among the members within a group are smaller than a predefined threshold ϵ_ϕ. By that, each G_i represents a proposition for the

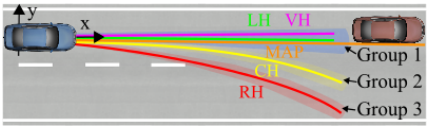


Fig. 9: Clustering the hypotheses regarding their angles.

driving direction. For example, Fig. 9 illustrates three groups resulted from five hypotheses. One condition here is that the rough road course from a navigation map (*MAP*) cannot form a group without further elements. The reason is that the used map is not accurate enough so that the vehicle can follow it blindly. Hence, the map is only a supporting source for the decision process but it is not a possible decision. Moreover, we add the group G_0 which reflects the proposition that all hypotheses are unreliable and should not be used. In such cases, the final result will be continuously tracked using only vehicle motion but without a "measurement update".

Secondly, the *frame of discernment* regarding *DST* is defined as $\Theta = \{G_0, G_1, \cdots, G_N\}$ where Θ is *exhaustive* (i.e., at least one G_i must occur) and its elements are *mutually exclusive* (i.e., at most one G_i can occur). Moreover, the power set Φ denotes the set of all possible subsets of Θ. Based on the hypotheses $h' \in H' = H \cup \{MAP\}$, we model the mass functions $m_{h'} : \Phi \rightarrow [0;1]$ with $m_{h'}(\emptyset) = 0$ and

$$m_{h'}(X) = \begin{cases} \frac{PR_{h'} \cdot R_{h'}}{K} & \text{if } \; h' \in X \\ \frac{1 - (PR_{h'} \cdot R_{h'})}{|\Phi| - K - 1} & \text{if } \; h' \notin X \end{cases}$$
$$\text{where } \; K = |\{Y \in \Phi | h' \in Y\}|$$

Thereby, $PR_{h'}$ represents the precision of the reliability estimator regarding R_h and serves as the third-order uncertainty. Our purpose of using $PR_{h'}$ is to penalize the classifiers with high misclassification rates. Afterwards, we combine all mass functions using the *DST*'s combination rule $\oplus$ to obtain the fused $m_F = m_0 \oplus m_1 \oplus .. \oplus m_{|H'|}$. Since $\oplus$ is commutative and associative [36], the order is irrelevant. Consequently, the belief and plausibility function are determined as

$$b_F(X) = \sum_{Y \subseteq X} m(Y) \text{ and } pl_F(X) = \sum_{Y \cap X \neq \emptyset} m(Y)$$

Finally, we choose the most probable proposition G_i by maximizing the average of the belief and plausibility functions with $\arg\max_{i \in [0;N]} \frac{b_F(G_i) + pl_F(G_i)}{2}$. When a group G_i with $i \neq 0$ is selected, a weighted-based combination of all members is performed. Thereby, the hypothesis MAP is excluded since it is inaccurate. Afterwards, this fused result is used for the measurement update of the Kalman filter.

C. Basic Approaches

Following, other fusion strategies are presented.

1) Baseline (BE): We employ a standard road estimation approach from [21] as the base line for the later evaluation.

2) Average Fusion (AVG): Hereby, all hypotheses are assigned with equal weights by the fusion.

3) Weight-based Fusion (WBF): Compared to *AVG*, *WBF* performs the fusion by using estimated R_h as weights.

4) Minimum (MIN): To prove that we can detect unreliable sources, *MIN* looks for the hypothesis with the smallest R_h.

5) Random (RAN): The purpose of *RAN* is to show that our fusion strategy is better than a random selection.

6) Winner-Takes-All (WTA): The hypothesis with the greatest R_h is selected without regarding the relation among all available hypotheses.

VI. Experimental Results

This section is devoted to present the experimental results by dividing them into three parts: feature selection, reliability estimation and fusion. Thereby, the database is built with many driving hours in and around Wolfsburg with the following distribution: 50% urban roads, 25% highways, 15% rural and 10% on- and exit-ramps (connections).

A. Evaluating Feature Selection

As described in Section IV-B , the observations of training data consist of three different sets of scenario features: sensor-related s, consensus τ and contextual information γ. Before applying feature selection, we first investigate the performance regarding each of these sets. For this purpose, we use *kNN* as the underlying classifier and show its performance measured by the *F-Score* in Table II. Here, the choice of the classifier

F	Highway				Connection			
	LH	RH	CH	VH	LH	RH	CH	VH
s_h	.827	**.879**	**.859**	**.964**	**.428**	**.481**	**.532**	**.922**
τ_h	.784	.827	.805	.963	.403	.363	.478	.910
γ_h	**.877**	.766	.842	.841	.362	.300	.434	.777

(a) Highways and connections.

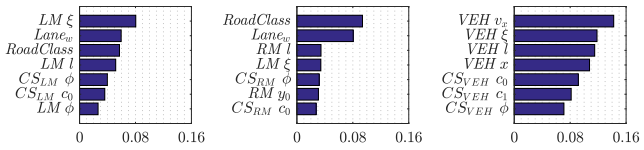
F	Rural				Urban			
	LH	RH	CH	VH	LH	RH	CH	VH
s_h	.820	**.864**	.854	**.927**	.630	.718	.714	**.956**
τ_h	**.856**	.848	.865	.902	**.718**	**.744**	**.736**	.934
γ_h	.848	.861	**.894**	.702	.666	.707	.696	.853

(b) Rural and urban roads.

TABLE II: Performance of classification with *kNN* regarding various feature groups.

is irrelevant since the focus is on the achieved performance of each feature set. This serves as a comparison to the works [7], [18], [23] that only use one of these sets. In detail, for highways and connections, Tab. IIa shows that using sensor-related information always has the highest F-Score. In contrast, there is not an obvious winner for rural roads as shown in Tab. IIb. Moreover, Tab. IIb shows that the consensus features play an important role for urban areas. This indicates that the consensus information might solve the problems occurring on urban roads, e.g., the ambiguity among the hypotheses, the absence of markings.

Using *Random Forests*, we perform feature selection to the set of all features X_h. For example, Fig. 10 presents the resulted set of the most relevant seven features for estimating R_{LH}, R_{CH} and R_{VH}. Fig. 10a illustrates that the existence value ξ of the left markings is the most important one for *LH*. Following this, all three sets of X'_{LH} are here represented with two features of each. Moreover, the most relevant features of *CH* also consist of information regarding the right markings



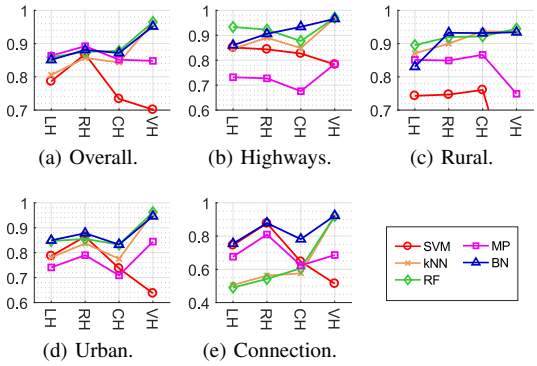
(a) LH. (b) CH. (c) VH.

Fig. 10: Most relevant features to estimate R_h of each h.

(Fig. 10b) since *CH* is built by using both markings. In contrast, the context information γ do not appear on the resulted top seven features of *VH* (Fig. 10c). Based on the feature selection's outcome, fifteen features are then selected to train the classifiers for each h.

B. Evaluating Reliability Estimation

As reliability estimation is performed by supervised learning approaches, this section measures the performance of the introduced classifiers. By that, we use the general terminology for the evaluation of classification. Since the reliable (positive) samples are overrepresented ($> 90\%$), we first down-sample them to get a balance between two classes. Otherwise, this class imbalance will lead to distorted results by just ignoring the minor class [34]. Moreover, we use the $F_{0.8} - Score$ to penalize the models by making misclassifications on the unreliable (negative) samples.



(a) Overall. (b) Highways. (c) Rural.

(d) Urban. (e) Connection.

Fig. 11: Classifiers' performance measured by $F_{0.8}$-*Score*.

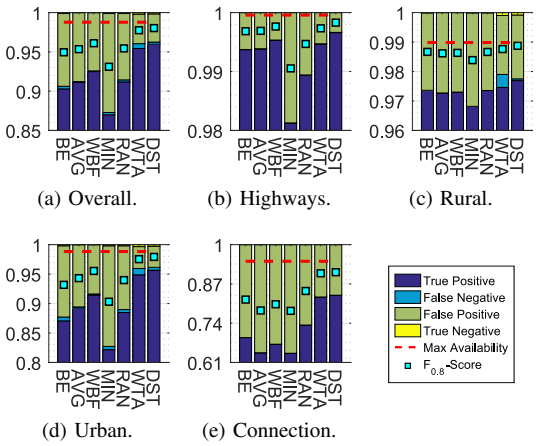
Accordingly, Fig. 11 presents the classifiers' performance regarding various scenarios. For highways, all models except *MP* achieve their best performance (Fig. 11b). The poor result of *MP* is caused by the data resampling so that the coordinates of many positive examples are not covered in the built model. Besides, *BN* and *RF* alternately achieve the highest F-Scores. In Fig. 11c, a similar behavior is obtained for rural roads with one difference: the worst model is *SVM*. Here, *MP* performs especially poorly by *VH* because *MP* is incapable of assessing reliability for dynamic hypothesis. For urban roads, Fig. 11d shows that *BN* outperforms other classifiers. This indicates that other models cannot handle complex situations in urban areas as well as *BN*. In connections, all classifiers have their worst performance (Fig. 11e) because of the challenging conditions such as the absence of the leading vehicles and narrow road curvatures. Eventually, *BN* is the most stable model over all scenarios and is then selected to support the fusion with its predicted reliabilities.

C. Evaluating Reliability-Aware Fusion

Finally, this section measures the performance of our entire system (Fig. 12). For this purpose, we compare the final output of the reliability-aware fusion module with the reference using the metric from Section IV-C. By that, the assessment is based on the Table III. Compared to Section VI-B, we do not resample the data here so that the results are related to the real road conditions.

		Final tracked output		
		$\Delta\alpha \leq 2°$	$\Delta\alpha > 2°$	**No output**
Any h Reliable	**Yes (Pos.)**	TP	FP	FN
	No (Neg.)	TP	FP	TN

TABLE III: Contingency table for the evaluation of the fusion



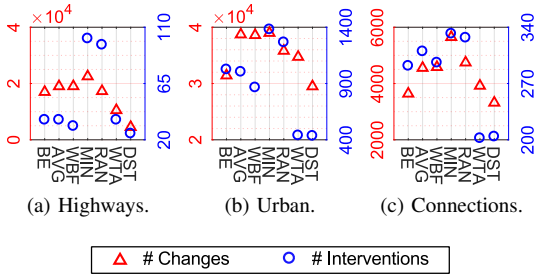
(a) Overall. (b) Highways. (c) Rural.

(d) Urban. (e) Connection.

Fig. 12: Final performance of fusion with different strategies and *BN* as reliability estimator.

In Fig. 12, the *maximum availability* (*MA*) denotes the amount of samples with at least one reliable hypothesis h. For highways and rural roads, Fig. 12b and 12c show that the performances of all strategies are near *MA* due to the good road conditions. As expected, the performance of all strategies decreases rapidly in urban areas (Fig. 12d) and reach the lowest values in connection (Fig. 12e). Overall, *DST* outperforms the rest since it also considers relations among the hypotheses. By ignoring this aspect, *WTA* selects only one h with the greatest R_h so that *WTA* takes second place. Although *AVG* and *WBF* combine all hypotheses, *WBF* achieves better performance by using R_h as weights. Moreover, the poorest result of *MIN* shows that the classifiers can identify the unreliable hypotheses. Compared to *BE* from [21], our fusion enables an increase of about 5 percentage points regarding the true positive rate (Fig. 12a).

To make strong case of *DST*, Fig. 13 compares the strategies regarding the number of the theoretical interventions [4] and the main hypothesis h_m changes between two consecutive frames. By that, h_m denotes the one with the greatest R_h among the selected hypotheses. In general, *DST* always has the smallest number of interventions. Moreover, *DST* switches least often between the hypotheses. These two aspects are important for a robust and stable road estimation with applications such as lane keeping function, etc.



(a) Highways. (b) Urban. (c) Connections.

Fig. 13: Further advantages of *DST*.

VII. Conclusion

This work presents a novel concept of incorporating reliability to support the multi-source fusion of the road estimation task. Thereby, the predicted reliabilities are obtained by using offline-trained classifiers. Based on Dempster-Shafer theory, the introduced fusion utilizes the reliability information for an optimal combination of the estimations. Compared to an average fusion, our concept can increase the availability of automated driving up to 8 percentage points. In future work, we will integrate further sources to receive additional redundant information, e.g., fish-eye cameras, grid-based road curve, etc. Moreover, we will investigate further classifiers and fusion strategies to improve the overall performance.

References

[1] G. L. Rogova and V. Nimier, "Reliability in information fusion: literature survey," in *7th Intl. Conf. On Information Fusion*, 2004, pp. 1158–1165.

[2] T. Brade, S. Zug, and J. Kaiser, "Validity-Based Failure Algebra for Distributed Sensor Systems," in *IEEE International Symposium on Reliable Distributed Systems*, 2013, pp. 143–152.

[3] J. Höbel, G. Jäger, S. Zug, and A. Wendemuth, "Towards a Sensor Failure-Dependent Performance Adaptation Using the Validity Concept," in *Computer Safety, Reliability, and Security*, S. Tonetta, E. Schoitsch, and F. Bitsch, Eds. Cham: Springer International Publishing, 2017, pp. 270–286.

[4] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to End Learning for Self-Driving Cars," *CoRR*, vol. abs/1604.07316, 2016.

[5] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[6] A. Bar Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: A survey," *Machine Vision and Applications*, vol. 25, no. 3, pp. 727–745, 2014.

[7] T. T. Nguyen, J. Spehr, M. Uhlemann, S. Zug, and R. Kruse, "Learning of lane information reliability for intelligent vehicles," in *IEEE Intl. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, 2016, pp. 142–147.

[8] T. T. Nguyen, J. Spehr, J. Xiong, M. Baum, S. Zug, and R. Kruse, "Online Reliability Assessment and Reliability-Aware Fusion for Ego-Lane Detection Using Influence Diagram and Bayes Filter," in *IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems*, 2017, pp. 7–14.

[9] P. Y. Shinzato, D. Gomes, and D. F. Wolf, "Road estimation with sparse 3D points from stereo data," in *IEEE Intelligent Transportation Systems Conf.*, 2014, pp. 1688–1693.

[10] L. Chen and H. Kong, "Lidar-histogram for fast road and obstacle detection," in *IEEE Intl. Conf. on Robotics and Automation*, 2017, pp. 1343–1348.

[11] Q. Li, L. Chen, M. Li, S. L. Shaw, and A. Nüchter, "A Sensor-Fusion Drivable-Region and Lane-Detection System for Autonomous Vehicle Navigation in Challenging Road Scenarios," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 2, pp. 540–555, 2014.

[12] H. F. Durrant-Whyte, "Sensor Models and Multisensor Integration," *Intl. Journal of Robotics Research*, vol. 7, no. 6, pp. 97–113, 1988.

[13] C. Gackstatter, P. Heinemann, S. Thomas, B. Rosenhahn, and G. Klinker, "Fusion of clothoid segments for a more accurate and updated prediction of the road geometry," in *13th Intl. IEEE Intelligent Transportation Systems Conference*, 2010, pp. 1691–1696.

[14] A. Abramov, C. Bayer, C. Heller, and C. Loy, "A flexible modeling approach for robust multi-lane road estimation," in *IEEE Intelligent Vehicles Symposium*, 2017, pp. 1386–1392.

[15] J. Thomas and R. Rojas, "Sensor-based road model estimation for autonomous driving," in *IEEE Intelligent Vehicles Symposium*, 2017, pp. 1764–1769.

[16] A. F. Garcia-Fernandez, M. Fatemi, and L. Svensson, "Bayesian Road Estimation Using Onboard Sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1676–1689, 2014.

[17] P. Wang, "Confidence as Higher-Order Uncertainty," in *Proc. of the 2nd International Symposium on Imprecise Probabilities and Their Applications*, 2001, pp. 352–361.

[18] F. Delmotte, L. Dubois, and P. Borne, "Context-dependent trust in data fusion within the possibility theory," in *1996 IEEE International Conference on Systems, Man and Cybernetics*, 1996, pp. 538–543.

[19] T. T. Nguyen, J. Spehr, H. Lin, and D. Lipinski, "Fused Raised Pavement Marker Detection Using 2D-Lidar and Mono Camera," in *IEEE Intl. Conf. on Intelligent Transportation Systems*, 2015, pp. 2346–2351.

[20] J.-F. Grandin and M. Marques, "Robust data fusion," in *3rd Intl. Conf.on Information Fusion*, 2000, pp. MOC3/3–MOC311 vol.1.

[21] D. Töpfer, J. Spehr, J. Effertz, and C. Stiller, "Efficient scene understanding for intelligent vehicles using a part-based road representation," in *IEEE Conf. on Intelligent Transportation Systems*, 2013, pp. 65–70.

[22] O. Hartmann, M. Gabb, R. Schweiger, and K. Dietmayer, "Towards autonomous self-assessment of digital maps," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2014, pp. 89–95.

[23] H. Guo, W. Shi, and Y. Deng, "Evaluating Sensor Reliability in Classification Problems Based on Evidence Theory," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 36, no. 5, pp. 970–981, 2006.

[24] M. Realpe, B. X. Vintimilla, and L. Vlacic, "A Fault Tolerant Perception system for autonomous vehicles," in *Proc. of the 35th Chinese Control Conference*, 2016, pp. 6531–6536.

[25] I. Bloch, A. Hunter, A. Ayoun, S. Benferhat, P. Besnard, L. Cholvy, R. Cooke, D. Dubois, and H. Fargier, "Fusion: general concepts and characteristics," *International Journal of Intelligent Systems*, vol. 16, pp. 1107–1134, 2001.

[26] A. Rechy Romero, P. V. Koerich Borges, A. Elfes, and A. Pfrunder, "Environment-aware sensor fusion for obstacle detection," in *Proc. of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2016, pp. 114–121.

[27] F.E. White, "A Model for Data Fusion," in *Proceedings of the First National Symposium on Sensor Fusion*, 1988.

[28] A. Polychronopoulos, A. Amditis, U. Scheunert, and T. Tatschke, "Revisiting JDL model for automotive safety applications: the PF2 functional model," in *Information Fusion*, 2006, pp. 1–7.

[29] K. Dietmayer, "Predicting of Machine Perception for Automated Driving," in *Autonomous Driving: Technical, Legal and Social Aspects*, M. Maurer and H. Winner, Eds. Springer Berlin, 2016, pp. 407–424.

[30] R. Kruse, C. Borgelt, C. Braune, S. Mostaghim, and M. Steinbrecher, *Computational Intelligence: A Methodological Introduction*, 2nd ed., ser. Texts in Computer Science. London: Springer London and Imprint: Springer, 2016.

[31] H. Liu and H. Motoda, *Computational Methods of Feature Selection*, ser. Data Mining and Knowledge Discovery. CRC Press, 2007.

[32] T. T. Nguyen, J. Spehr, J. Xiong, M. Baum, S. Zug, and R. Kruse, "A Survey of Performance Measures to Evaluate Ego-Lane Estimation and a Novel Sensor-Independent Measure Along with Its Applications," in *IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems*, 2017, pp. 239–246.

[33] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[34] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. Springer, 2006.

[35] G. Shafer, *A mathematical theory of evidence*. Princeton NJ u.a.: Princeton Univ. Press, 1976.

[36] F. Ye, J. Chen, Y. Li, and J. Kang, "Decision-Making Algorithm for Multisensor Fusion Based on Grey Relation and DS Evidence Theory," *Journal of Sensors*, vol. 2016, no. 3, pp. 1–11, 2016.