Evidence Based Analysis of Internal Conflicts Caused by Disparity Inaccuracy

Thorsten Ike Baerbel Grabe Florian Knigge Michael Hoetter
Institute of Innovation-Transfer
University of Applied Sciences and Arts Hannover
Hannover, Germany.
{thorsten.iike; baerbel.grabe; florian.knigge; michael.hoetter}@fh-hannover.de

Abstract – For future driver assistance and safety systems highly reliable and comprehensive maps of the car’s environment are needed. The map quality depends highly on the accuracy and the interpretation of a given set of measurements. This contribution introduces two metrics derived from Dempster-Shafer theory. By means of these metrics two different inverse sensor models are compared. This comparison is done for two given sets of data from a stereo video sensor with pixel and subpixel accuracy. The introduced metrics reflect the abilities of the inverse sensor models to build reliable and comprehensive maps. Furthermore they show the connection between these abilities, the inverse sensor model parameter and the measurement accuracy.

Keywords: Data fusion, inverse sensor models, internal conflict, evidence theory.

1 Introduction

One of the major challenges for the development of future driver assistant and safety systems is a complete and highly reliable description of the car’s environment. In particular for systems with the ability of autonomous or semi-autonomous vehicle control intervention, maybe only one malfunction in 100 vehicle years or even less could be seen as the acceptable limit of their trustworthiness [6].

Complete knowledge of the car’s environment requires that all objects on one hand and all free spaces on the other hand are determined. The required high reliability can only be reached if the fusion of these complementary environment descriptions does not lead to inconsistencies or, more precisely, conflicts for certain areas. In this paper we compare two different inverse sensor models (Figure 1) which are used for the interpretation of two sets of measurements with different accuracy. Furthermore we show the interconnection between model parameters, conflicts, measurement accuracy and the knowledge about the vehicle’s surroundings.

Conflicts arise if different sensors determine contrary states for a part of the vehicle’s surroundings, e.g. an area is determined as occupied and free. If contrary states are derived from a single sensor, we call it an internal conflict. These conflicts lead to inappropriate maps [1] and hence, they corrupt the reliability of safety systems.

Occupancy grids, introduced by [2] (section 2), represent the observed vehicle’s environment as a finite number of equal sized cells. Depending on the available sensor data and the used inverse sensor models, each cell is given a probability of being occupied or empty [3][9][10]. Section 3 describes our data fusion concept based on the evidence theory, also known as Dempster-Shafer Theory (DST), established by [7]. We define two evidence based metrics which describe the capability of different inverse sensor models for a comprehensively description of the vehicle’s surroundings. Section 4 presents two different inverse sensor models and points out that the interpretation of range measurements is not unique and lead to varying internal conflicts. Our experimental results in section 5 demonstrate the connection between the different inverse sensor model parameters, the accuracy of the measurements and the evidence based metrics.

Figure 1: Interrelationship between sensor model, inverse sensor model and occupancy grid according to [2]. Left: Scene in front of the vehicle described by a set of attributes W. Middle: On measurements mapped attributes using sensor models. Right: From inverse sensor models derived occupancy grid, high and bright illustrated cells represent a high occupancy probability. The white lines show detected lane markings.
2 Occupancy Grid Representation

Occupancy grids are a quantised representation of the vehicle’s surroundings. In our approach the area in front of the car is divided into a finite number of cells with equal size of 0.2m · 0.2m. The i-th cell in x- and the j-th cell in z-direction is denoted as C_{i,j}. The probability of a cell C_{i,j} being occupied or empty is defined as P_{occ}(C_{i,j}) and P_{emp}(C_{i,j}) respectively. At this point we should carefully remark that these probabilities are derived from independent inverse object and free space models. Hence, we have P_{emp}(C_{i,j}) = 1 - P_{occ}(C_{i,j}), different from approaches like [3][10] the free space probability does not depend on detected objects.

3 Data Fusion Using Dempster-Shafer Theory

The Dempster-Shafer theory is based on a set of exhaustive and exclusive states, called the frame of discernment Θ. In this contribution, the frame of discernment includes all possible states of a cell C, Θ = {\{(occ)uped\}, \{(emp)ty\}}. With all possible states the power set of Θ is defined as 2^Θ = {∅, {occ}, {emp}, Θ} which is a partially ordered set where the supersets (here Θ) include finer, disjunctively combined sets. Similar to the probability density in classic probability theory, a mass function assigns support for a subset of 2^Θ with: m : 2^Θ → [0, 1] and

\[ m(∅) = 0, \sum_{A⊂Θ} m(A) = 1. \]  

Contrary to the classic probability theory, where support for a state (e.g. P(occ) = 0.8) implies support to its complement (P(emp) = 0.2), equation (1) allows to give support to Θ, which represents ignorance. 

Dempster’s rule of combination combines two independent mass functions m_1 and m_2 over the frame of discernment:

\[ m(X) = \begin{cases} \frac{1}{1-m'(<0)} m'(X) & \text{for } X ≠ ∅ \\ 0 & \text{if } X = ∅ \end{cases} \]  

m'(X) = \sum_{A_i∩B_j=X} m_1(A_i) · m_2(B_j)  

The result is called the orthogonal sum and is written as m = m_1 ⊕ m_2. With repeatedly applying equation (2) an arbitrary number of mass functions can be combined. A specific relevance comes with the renormalisation term in equation (2). It is the conflict mass within the weight of conflict which is defined as:

\[ con = -\log(1 - κ) \] with

\[ κ = m'(∅) = \sum_{A_i∩B_j=∅} m_1(A_i) · m_2(B_j) \]  

With raising inconsistency of the mass functions m_1 and m_2, con goes to infinity. Thus con = ∞ means that equation (2) is in an undefined state. Furthermore the degree of belief and the degree of plausibility as defined in equation (4) and (5) are the minimal and maximal support for a given state. They are interpreted as lower and upper boundaries of an interval whereas the interval width describes the remaining uncertainty of the fused data.

\[ Bel(X) = \sum_{A⊂X} m(A) \]  

\[ Pl(X) = \sum_{A∩X≠∅} m(A) \]  

For the used frame of discernment in this contribution the interval width is equal to the remaining ignorance Θ.

The DST has often been criticised producing counterintuitive results in some cases. The most famous example is from [11] and argues as follows:

A patient P is examined by two doctors A and B. A’s diagnosis is that P has either meningitis, with probability 0.99, or brain tumor, with probability 0.01. B agrees with A that the probability of brain tumor is 0.01, but believes that it is the probability of concussion rather than meningitis that is 0.99. Applying Dempster’s rule leads to the conclusion that the belief that P has brain tumor 1.0 - a conclusion that is clearly counterintuitive.

As pointed out in [4] this example has some weaknesses. We follow this argumentation and add another point of view within the context of this paper. At first the doctors examine the patient. Here, the doctor is a kind of sensor. The examination produces uncertain data. At next, the doctors interpret the data to diagnose the possible illnesses (this can be considered as a sensor model). Applying Dempster’s rule gives the mentioned result. But for a complete interpretation one has to consider another important result, the weight of conflict. The high value of con in this example shows that, if the sensor models are true (the examinations are reliable), the interpretation, and therefore one of the inverse sensor models (the diagnose of one doctor), is false.

For evaluating the inverse sensor models in section 4 two ignorance based metrics are defined:

\[ Π_m = \frac{1}{N} \sum_{C_{i,j}} 1 - m_{C_{i,j}}(Θ) \forall C_{i,j} \]  

\[ Π_d = \frac{1}{N} \sum_{C_{i,j}} 1 - m_{C_{i,j}}(Θ) \forall C_{i,j} : con_{C_{i,j}} ≤ t_σ \]  

where N is the total number of cells and t_σ is used as a threshold to ignore conflicts evolved from noise.

Π_m summarises the knowledge of an inverse sensor model about the vehicle’s environment. In contrast, the Π_d metric summarises only conflict free knowledge.
of the vehicle’s surrounding. The most comprehensive inverse sensor model leads to a high value of $\Pi_m$, and the most reliable inverse sensor model leads to a high value of $\Pi_d$. Hence, $\Pi_d = \Pi_m = 1$ is the best possible inverse sensor model. $\Pi_m$ and $\Pi_d$ are highly dependent on the inverse sensor model parameter shown in section 4 and they are used in this paper to clarify which inverse sensor model for given sets of data gives the most reliable interpretation. In addition these metrics enlighten the influence of the measurement accuracy on the different inverse sensor models.

4 Inverse Sensor Models

In this section we briefly introduce our stereo video sensor. This is followed by the presentation of the different inverse sensor models. The requirements of the inverse sensor models are twofold. On one hand detailed inverse models are preferable for getting most information out of the measurements, but on the other hand those models are often too specific to detect all kind of objects in front of a car (Figure 2). Hence, the question is: How much inverse sensor model specification is needed for an appropriate representation of the scene, sensed by a stereo video sensor, without loss of generality? We compare two inverse sensor models with a different grade of specification and show in section 5 the connection between their parameters and the ability for a reliable representation of the car’s environment.

4.1 Sensor setup

Our video sensor is a calibrated stereo vision system. The images from the two cameras are rectified [5]. With the resulting standard stereo geometry we perform a disparity estimation which uses the zero-mean sum of squared differences to calculate the point correspondences with pixel accuracy. For the subpixel disparity estimation the expectation of the zero-mean sum of squared difference distribution over the image coordinates is used [8]. Based on the point correspondences 3d-dimensional points are reconstructed by triangulation (Figure 3). The uncertainty of an estimated 3d-point is given by its covariance matrix $\Sigma_{\hat{X}}$: 

$$\Sigma_{\hat{X}} = \Sigma_{\hat{X},p} + \Sigma_{\hat{X},d}, \quad (8)$$

where $\Sigma_{\hat{X},p}$ represents the uncertainty of the 3d-point caused by parameter uncertainties estimated in the sensor calibration process and $\Sigma_{\hat{X},d}$ the uncertainties from the estimation of point correspondences between the stereo image pair, respectively. Let $L$ be the number of all reconstructed 3d-points in the observed area. Then the set of data is given by:

$$\hat{X}_k, \Sigma_{\hat{X},k} \quad k \in L. \quad (9)$$

Figure 3: Birdseye view of the reconstructed 3d-points from the stereo video sensor. Left: 3d-points derived from subpixel precise disparity estimation. Right: 3d-points derived from pixel precise disparity estimation.

4.2 Inverse sensor model for single 3d-points

As first inverse sensor model we take the simplest approach: the intention is to use every single measurement as a new independent piece of evidence getting a guess about the maximum information about the car’s environment at hand. Every single measurement $\hat{X}_k$ is considered as an unbiased estimation of the ’true’ value $X_k$ and assumed to be Gaussian distributed. Then, the probability density function of $\hat{X}_k$ is defined as:

$$p(\hat{X}_k|X_k, \Sigma_{\hat{X},k}) \quad (10)$$
Two thresholds separate the set of \( L \) measurements in \( F \) measurements representing free space (height coordinate of \( \hat{X}_k \leq t_1 \)) and \( O \) measurements interpreted as objects (height coordinate of \( \hat{X}_k \geq t_h \)). Every single point from these two separated sets gives a new piece of evidence to all cells within the 3\( \sigma \) limits of \( \Sigma_{X_k} \) being empty,

\[
m_{C_{i,j}}^{k}\left(\{\text{emp}\}\right) = \int_{C_{i,j}} p(\hat{X}_k|\hat{X}_k,\Sigma_{X_k}) \mid k \in F \quad (11)
\]
or occupied, respectively. For every affected cell the ignorance is stated as

\[
m_{C_{i,j}}^{k}(\Theta) = 1 - m_{C_{i,j}}^{k}\left(\{\text{emp}\}\right) \mid k \in F \quad \text{and}
\]

\[
m_{C_{i,j}}^{k}(\Theta) = 1 - m_{C_{i,j}}^{k}\left(\{\text{occ}\}\right) \mid k \in O. \quad (12)
\]

For fast computing over the final grid Dempster’s rule is accessed separately for all object mass functions and for all free space mass functions, yielding to a simple combination rule for the separated sets with low computational costs:

\[
m_{C_{i,j}}\left(\{\text{emp}\}\right) = 1 - \prod_{k}(1 - m_{C_{i,j}}^{k}(\{\text{emp}\})) \mid k \in \{1,2\}
\]

\[
m_{C_{i,j}}(\{\text{occ}\}) = 1 - \prod_{k}(1 - m_{C_{i,j}}^{k}(\{\text{occ}\})) \mid k \in \{1,2\}
\]

Here, equation (2) is accessed with the results of equations (11)-(14) and (2). For the occupied cells, the covariance matrix in equation (11) is the sum of the covariance matrices of the highest and lowest point.

4.3 Inverse sensor model for 3d-point cluster

For the second inverse sensor model we adopt the concept of a \textit{terrain labeling function} according to [9]. The main idea is that a free space can be represented through a single measurement, but an object has to be measured at least twice with a significant height difference (Figure 4). Hence, in a first step a cell \( C_{i,j} \) is

- **Empty** if at least one point is within the \( \epsilon \)-range around the center of cell \( C_{i,j} \) with an absolute height coordinate below \( t_1 \). Again, if more than one point is found, the most reliable measurement is stored.

- **Unknown** if no point exists within the \( \epsilon \)-range around the center of cell \( C_{i,j} \).

The stored measurements which define the cell states are used in the second step to calculate the cell probabilities being occupied or empty with equations (11)-(14) and (2). For the occupied cells, the covariance matrix in equation (11) is the sum of the covariance matrices of the highest and lowest point.

5 Experimental Results

As mentioned before, for the most comprehensive (all of the car’s environment is known) and most reliable (no conflicts) inverse sensor model \( \Pi_d = \Pi_m = 1 \) holds. Figures 5-12 show the experimental results for the different inverse sensor models from section 4. Figures 5-8 are demonstrating the results based on the measurements with pixel accuracy and Figures 9-12 the results derived from interpreting the data with subpixel accuracy.

Every figure reflects the dependency of \( \Pi_m \) and \( \Pi_d \), respectively, on the varying parameters \( t_h \), \( t_{md} \) and \( t_l \) and therefore their ability for a comprehensive and reliable mapping of the car’s environment.

The inverse model for the single 3d-points holds the highest \( \Pi_m \) which is sensitive for the free space threshold represented by the parameter \( t_l \) (Figure 5), but almost independent of the object description parameter \( t_h \). \( \Pi_d \) reaches its maximum at \( t_h = 0.9m \) (Figure 6). This points out that this inverse model generates the densest but least reliable map. Nearly 50\% of the cells are corrupted by conflicts if a parameter \( t_h > 0.6m \) is used. The increasing \( \Pi_d \) for \( t_h > 0.6m \) is due to the fact that many objects are no longer detected.

The results using this inverse sensor model with the 3d-points based on the subpixel precise disparities are almost identical (Figure 9 and Figure 10). Hence, in combination with the inverse sensor model for the single 3d-points, a disparity estimation with subpixel accuracy does not lead to a more reliable or more comprehensive map.

Using the inverse model based on point cluster leads to a highly reliable but incomplete map for the 3d-points derived from pixel precise disparities. Only in the case that objects smaller than 0.2m are accepted, noticeable conflicts appear (Figure 7 and Figure 8). The low values for \( \Pi_m \) and \( \Pi_d \) reflect the dropped information since at least two points are necessary to define an occupied cell.

As a result of using the subpixel precise data, \( \Pi_m \) and \( \Pi_d \) are increasing by a factor of three (Figure 11.
Figure 5: Experimental results of $\Pi_m$ for the inverse sensor model based on single 3d-points derived from disparities with pixel precision

Figure 6: Experimental results of $\Pi_d$ for the inverse sensor model based on single 3d-points derived from disparities with pixel precision

Figure 7: Experimental results of $\Pi_m$ for the inverse sensor model for 3d-point cluster derived from disparities with pixel precision

Figure 8: Experimental results of $\Pi_d$ for the inverse sensor model based on single 3d-points derived from disparities with subpixel precision

Figure 9: Experimental results of $\Pi_m$ for the inverse sensor model based on single 3d-points derived from disparities with subpixel precision

Figure 10: Experimental results of $\Pi_d$ for the inverse sensor model based on single 3d-points derived from disparities with subpixel precision
Figure 11: Experimental results of $\Pi_m$ for the inverse sensor model for 3d-point cluster derived from disparities with subpixel precision.

Figure 12: Experimental results of $\Pi_d$ for the inverse sensor model for 3d-point cluster derived from disparities with subpixel precision.

and Figure 12). The subpixel disparities spread the 3d-points over a wider area of the occupancy grid (Figure 3). Hence, more grid cells are fulfilling the condition being occupied or empty.

6 Conclusions

In this paper we have compared different interpretations for two data sets from a stereo video sensor. For this purpose we used the Dempster-Shafer theory and introduced two DST based metrics, $\Pi_m$ and $\Pi_d$, which reflect the ability of the inverse sensor models to build comprehensive and reliable maps for measurements with different accuracy.

It has been shown that using the inverse sensor model for the single 3d-points, a disparity estimation with subpixel accuracy does not lead to a more reliable or more comprehensive map. Using the inverse sensor model based on point cluster with the subpixel precise disparities increases $\Pi_m$ and $\Pi_d$ by a factor of three.

Furthermore we have shown that these metrics could be used as optimisation criteria for these models. In future we will expand our investigations to different grid resolutions and adopt our approach to the optimisation of a multisensor network.

References