



HAL
open science

Heterogeneous sensor data fusion for multiple object association using belief functions

H. Laghmara, T. Laurain, C. Cudel, J.P. Lauffenburger

► **To cite this version:**

H. Laghmara, T. Laurain, C. Cudel, J.P. Lauffenburger. Heterogeneous sensor data fusion for multiple object association using belief functions. *Information Fusion*, 2020, 57, pp.44 - 58. 10.1016/j.inffus.2019.11.002 . hal-03489084

HAL Id: hal-03489084

<https://hal.science/hal-03489084>

Submitted on 21 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Heterogeneous Sensor Data Fusion for Multiple Object Association using Belief Functions

H. Laghmara^{a,*}, T. Laurain^a, C. Cudel^a, J. P. Lauffenburger^a

^a *Université de Haute-Alsace, IRIMAS-EA7499, 12 rue des frères Lumière, F-68093 Mulhouse, France*

Abstract

In this study, the object association issue is tackled in order to ensure a correct affiliation of perceived objects with known ones. The proposed approach is based on the evidence theory and includes multiple object features in order to manage pairing issues in a complex environment. Two heterogeneous information sources are built based on kinematic features related to the objects: their position and size on one hand and their direction of motion on the other hand. A study on the estimation of the belief expressed by these independent sources is performed. The multiple features are managed through a hierarchical fusion which includes two levels of combination. The first level is a pairwise combination, fusing position and orientation data of each pair of objects and the second one processes sequentially the previously combined information over all possible associations. This paper also investigates the effectiveness of the association according to different combination operators at both levels. The performance of the proposed approach is demonstrated in the Intelligent Transportation Systems context for which environmental perception is crucial. The validation exploits a large amount of real data issued from a camera and a 3D LiDAR from the KITTI database.

Keywords: Multiple object association, Evidence Theory, Multi-attribute

*Corresponding author

Email addresses: hind.laghmara@uha.fr (H. Laghmara), thomas.laurain@uha.fr (T. Laurain), christophe.cudel@uha.fr (C. Cudel), jean-philippe.lauffenburger@uha.fr (J. P. Lauffenburger)

1. Introduction

Multiple Object Association (MOA) is a crucial function in Multi Target Tracking (MTT) consisting in matching perceived objects (targets) to known ones (tracks). This can be done in a multi-sensor configuration in order to
5 match, at a time, the detections from several sensors. Another solution is to consider data from one sensor over time leading then to a temporal association problem. For Intelligent Transportation Systems (ITS) applications, semantic scene interpretation is often based on sensors which might typically provide uncertain, imprecise or incorrect data. This results in mis-detections or ambiguous
10 information about the true identity or location of an object. Multi-sensor data fusion offers a solution to this issue as it seeks to combine data to perform inferences that may not be possible from a single sensor [1].

Several MOA methods exist: Nearest Neighbour, (Joint) Probability Data Association Filter ((J)PDAF) or Multi-Hypothesis Tracking (MHT), all defined
15 on the Bayesian frame [2, 3]. However, in the presence of uncertainties, one can prefer the theory of belief functions or evidence theory. It is particularly convenient for uncertainty modeling and propagation in the combination of partially reliable information. Introduced by Dempster and Shafer [4, 5] as a generalization of the Bayesian framework, it has later been used by Smets in his formalization of the Transferable Belief Model (TBM) [6]. The use of this theory in MOA
20 has brought many advantages compared to the probabilistic methods such as the management of object appearances and disappearances, or its average complexity. In [7], Gruyer *et al.* highlight the interest of belief function-based MOA (BF-MOA) w.r.t. the bayesian approaches.

25 Hereafter, the authors first describe the major contributions in BF-MOA with a focus on the latest ones. The contributions of this paper are given afterwards.

1.1. Related Work

Rombaut was first to model MOA with belief functions by defining many
30 object attributes (numerical, symbolic and logical) [8, 9]. In these works, estimation models of the attribute mass functions widely used in further research have also been introduced. Based on Rombaut’s formalism, Royère *et al.* [10] established generalized combination rules considering the *extended open world* model (*ew*). In *ew*, track appearances and disappearances are possible. Improving this work, Mourllion *et al.* [11] developed two pairing solutions to avoid
35 suspicious local associations when evidence conflicts. [11] was also at the origin of belief function-based MHT algorithms which have largely inspired Gruyer *et al.* [7] for road context evaluation. Nevertheless, their solutions suffer either from high complexity and not dealing with the conflict, or from inaccuracy in
40 the generated track numbers. Later, Mercier *et al.* [12] proposed a decisional process in the TBM framework aiming at maximizing the pignistic probabilities in a global way. In this ITS-related study, a fusion of the distances and angles between the ego-vehicle and the detected objects is performed.

Denceux *et al.* and Daniel *et al.* [13, 14] highlighted that Mercier’s solution
45 is computationally expensive and intractable for real-time execution. Moreover, it can also conclude to suspicious associations due to the global optimization. In the last years, Daniel *et al.* [14, 15] enhanced the process by suggesting two pairing algorithms which locally solve the association problem and largely reduce the computational complexity.

50 In most of the studies referenced above, the association of multiple perceived objects in a scene is based on the position information [7, 11, 12]. When available, additional object features like its class or velocity have been considered as pieces of evidence to be gathered. Ristic and Smets [16], for instance, use the class information with the assumption that objects from the same class could
55 be associated. **Inspired by these works, Rekik *et al.* [17] followed the same methodology to construct objects in a scene starting from partially occluded detections. The association part is done thanks to a modified Hungarian al-**

gorithm. Considering that identical class objects can be matched is a strong assumption. Especially in road scenes, it is common to have objects of the same class (pedestrians or cars on highway) in close positions or to have a number of
60 classes close to the number of objects.

Dencœux *et al.* [13] solved the MOA by maximizing the plausibility of all possible relations between tracks and targets through a linear assignment problem taking account of three object features: velocity, class and position. The main
65 advantage is that the resulting computational cost is polynomial. However, the constructed relations are exclusive to the number of tracks and targets and the decision criteria has the inconvenience of being optimistic.

Chavez-Garcia and Aycard [18] developed a real-time multi-sensor fusion system devoted to road scene perception exploiting the evidence theory. By
70 adding the object classification to the their kinematic features provided by three sensors (LiDAR, radar and camera) and by fusing them with Yager's rule (see Section 3.4), they showed improved detection results. To achieve good performance, an important a priori knowledge of the object classes characteristics is necessary to define the *bbas*. In [19], Hachour *et al.* introduced a matching algo-
75 rithm based on the Generalized Bayes' Theorem and evaluated its performance on synthetic data (simulated tracking scenarios). Nevertheless, only mono-scan and mono-sensor configurations are considered and the complexity is higher than the reference approach from Dencœux [13]. A multi-sensor tracking method has been recently proposed by Dan *et al.* [20]. In their work, an association step
80 very similar to the one presented in [21], is done thanks to an orthogonal sum to link local tracks with fused tracks. Nevertheless, the pairing results are highly related to the multi-sensor tracking model. In the proposed paper, the focus is put only on the object association step. It appears from this review that track association in cluttered environment should consider heterogeneous object fea-
85 tures in order to be effective. Those features have to be selected depending on the application considered.

1.2. Contributions

As shown previously and to the best of the authors' knowledge, the literature is mainly dedicated to enhance the final step of the association process, i.e. the decision-making, or to suggest formalizations dealing with the computational
90 complexity. The aggregation of potentially discordant data and the resulting conflict has not yet been exploited in the aim of dealing with suspicious or ambiguous associations. In fact, numerous contributions are providing improved pairing decision criteria and algorithms (joint pignistic probability in [12], local
95 pignistic probability in [14], contour function in [13], etc.). Only few of them consider the conflict raised during mass combination [7, 11]. Moreover, none of them discuss the choice of the fusion rule and use generally Dempster's rule, even if the pairwise pairings are based on the combined data. It is then obvious that the fusion step of full, imperfect or heterogeneous information plays a
100 key role in the association scheme. In practical applications like real-time environment perception, it is common to deal with partial, imprecise, uncertain or missing information. In this context, complex situations can lead to incomplete or conflicting data given by the sensors. That is why, this work focuses on the choice of the combination operators with respect to the association performance.
105 A main interest is put on the conflict management raised during combination.

The idea behind this paper is to define, for ITS-related applications, a robust object association framework dealing with multiple sources and heterogeneous data. The focus is put on an adequate modeling of the information according to their distinct nature. We propose in this paper a review of the data models
110 suitable for this objective. A thorough study of the combination of the data delivered by heterogeneous sources is presented to guarantee a robust result regardless of the matching algorithm. We believe that if an information is discriminant enough, the solution issued from the combination algorithm must be trivial. Therefore, we investigate another additional source to the object position, expressing the motion direction of dynamic objects. Object features like
115 their class already exploited in the literature and with limited impact for road

scene perception require classification methods unintended in this application. The direction of motion of the perceived objects can, in complex scenarios like intersections or highway driving, help to discriminate those objects.

120 The objectives of this article are multiple : determine the appropriate source model for the retained object attributes (cf. Section 4.2), propose a robust multiple feature fusion strategy and choose the best combination rule (cf. Section 4.1) and finally evaluate the approach on a large amount of realistic driving scene data (cf. Section 5).

125 The aim of this paper is to present an original strategy to improve and extend a camera-based MOA algorithm [21] with the use of LiDAR data. The knowledge on which the association is applied is derived on the one hand from the object's position and size obtained from images of the scene. On the other hand, the object direction of motion are extracted from 3D Lidar point clouds.

130 Both data represent two different similarity measures for the construction of two credal sources. A multiple feature fusion strategy is applied to manage these heterogeneous sources and to guarantee the best outcome through different combination rules. A two step hierarchical fusion strategy is introduced: the first one handles, through a multisensor and temporal fusion, data from

135 heterogeneous sources expressing a local knowledge of each candidate hypothesis. It combines position and orientation pairwise mass functions and provides the belief in the association of two specific objects. The second one allows to raise the ambiguity by expressing a global knowledge based on an aggregated belief w.r.t. all available sources. It is a sequential fusion which gathers the

140 knowledge expressed in the first level of combination in order to define the pairing confidence between resp. a given target/track and resp. all tracks/targets. To assess the feasibility of the fusion framework in the application context of autonomous vehicles, the KITTI database [22] is used. A quantitative analysis performed on an important amount of data containing objects of different

145 nature is done. To the authors knowledge, such an extensive validation (considering several hundreds of frames and pairings) of a belief function-based object association solution in real conditions is unique.

The article is organized as follows: the application context and problem description are given in Section 2. Some fundamental concepts in evidence theory and their application to MOA are presented in Section 3. The modeling
150 of the object features and the multiple feature fusion strategy are then described in Section 4 to be later tested on experimental data in Section 5. Finally, Section 6 concludes this paper.

2. MOA for Intelligent Vehicles

155 2.1. Object Perception in Cluttered Environments

In the context of Intelligent Vehicles (IV), the crucial point is to ensure the safety of road users. In order to achieve such an objective, it is mandatory for an IV to obtain an accurate and robust representation of the environment near the vehicle. The main aim of a perception system is then to detect the surround-
160 ing environment (static and moving objects, navigable area, etc.). Usually, perception starts by collecting data from multiple and heterogeneous sensors (cameras, radars, LiDARs, etc.) mounted on the ego-vehicle. As none of these hardware solutions provide perfect data, there is a need to fuse the information collected. Moreover, objects in the scene can be totally/partially occluded, external conditions can vary, etc. Thus, providing tools able to model/deal with
165 data uncertainties, imprecision, incompleteness or conflicting information is of utmost interest. In the context of a complex surrounding environment (urban traffic, pedestrians...), the perception system should be robust enough to handle several objects to be detected and associated [23]. In this paper, the input of the
170 MOA algorithm is based on camera images. As already mentioned, from these 2D data, it is possible to determine the position and the size of the objects. In addition, the use of a LiDAR provides 3D data from the point cloud (see Figure 1). The LiDAR information can be used to improve the MOA algorithm by including, in addition to the position from the camera, the orientation of the
175 object. Indeed, as one can see on Figure 1 taken from the KITTI benchmark [22], the camera-based image can be naturally enriched with LiDAR informa-

tion. The perception of objects in cluttered environment, and the temporal association of perceived objects, is here realized through the fusion of measures from these two heterogeneous sensors.

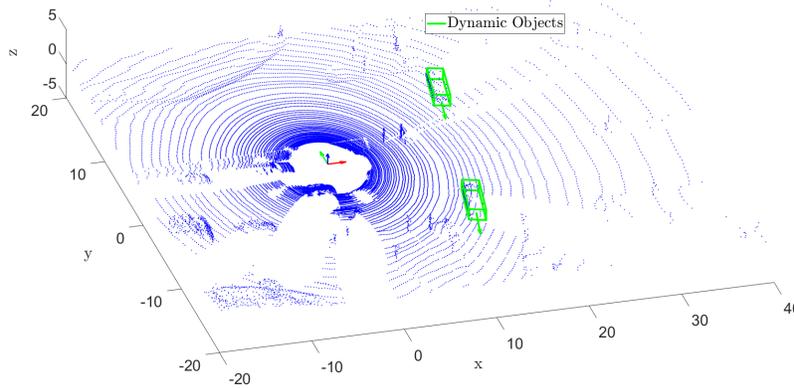


Figure 1: Example on a typical scene from the KITTI benchmark (Top) and the corresponding LiDAR Data (Bottom).

180 *2.2. Problem Description*

Let us define respectively a target/track as X_i/Y_j . Basically, the pairing process performed during data association between multiple targets and tracks aims at associating a target X_i to a track Y_j . In highly changing environments, the process is challenging due to the variable object numbers as well as their possible appearance and disappearance over time. The association process is bi-directional (see Fig. 2), i.e. two cases of pairing are distinguished: a target-to-track ($X \rightarrow Y$) and a track-to-target ($Y \rightarrow X$) matching.

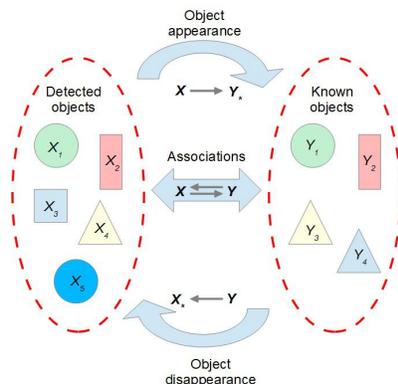


Figure 2: Multi-Object association from [24].

Two sets Θ_{X_i} and Θ_{Y_j} gathering respectively the M tracks and N targets are defined:

$$\Theta_{X_i} = \{Y_1, Y_2, \dots, Y_M, Y_*\}, \quad (1)$$

$$\Theta_{Y_j} = \{X_1, X_2, \dots, X_N, X_*\}, \quad (2)$$

190 with $i \in I = \{1 \dots N\}$, $j \in J = \{1 \dots M\}$. A track appearance is described by including the solution Y_* to the track set Θ_{X_i} whereas a track disappearance is defined by the proposition X_* in the target set Θ_{Y_j} . The association task to be treated remains as answering the question, considering a pair of objects at a time k : “are these objects related?” and two possible solutions are whether
 195 “yes” or “no”.

3. Representation of Evidence

This section is dedicated to some fundamental tools of the Dempster-Shafer framework for evidence modeling and decision-making under uncertainty. To complete this synthesis, the reader may refer to the references given below as well as to the seminal book of Dempster [5] and to Smets’s work regarding the
 200 TBM [6].

3.1. Basics of Belief Functions

The mathematical theory of evidence has been introduced by Shafer in order to ensure an adequate representation of uncertainty. Shafer's work is often called theory of evidence because it deals with weights of evidence and numerical degrees of support based upon evidence. A *frame of discernment* Θ is firstly defined as a set of discrete states (hypotheses H_j):

$$\Theta = \bigcup_{j=1}^k \{H_j\} \text{ with } \forall i \neq j, H_i \cap H_j = \emptyset. \quad (3)$$

k is the number of hypotheses and Θ describes all known exclusive solutions to the tackled problem. All possible combinations of singletons H_j of Θ define the power set or *referential frame* 2^Θ such that:

$$2^\Theta = \{\emptyset, \{H_1\}, \dots, \{H_j\}, \dots, \{H_1, H_2, H_3\}, \dots, \Theta\}. \quad (4)$$

When modeling knowledge, pieces of evidence can be set on one or several elements of 2^Θ . This evidence from a source is represented by a *mass function* $m(\cdot)$ defined as a mapping from the power set 2^Θ to $[0, 1]$:

$$\sum_{A \in 2^\Theta} m(A) = 1, \quad (5)$$

with $m(A)$ the degree of belief supporting proposition $A \in 2^\Theta$ and that can not be committed to any more specific proposition of the referential subset. A is known as a focal set. $m(A)$ defines the belief that A is the solution to the raised question/problem.

Based on these definitions, the aim when addressing a real problem is to construct an evidential model, i.e. a *basic belief assignment* (*bba*) for any source expressing knowledge on the problem to be solved.

3.2. Belief Assignment Models

In order to define the source knowledge, an appropriate *bba* should be estimated. In the literature, several models are available and the chosen one oftenly

depends on the nature of data and the application in concern. Some common models are described and discussed here.

A *simple bba* has at most two focal sets [4]. This model is used when a source S_j can only support one hypothesis $A \in 2^\Theta$ and attributes a partial knowledge to this solution:

$$\begin{cases} m_j(A) = s, \\ m_j(B) = 1 - s, \\ m_j(C) = 0, \forall C \in 2^\Theta, C \neq A, C \neq B, \end{cases} \quad (6)$$

with $s \in [0, 1]$. When $B = \Theta$, this model is known as a *simple support function*. With $B = \bar{A}$, a *complementary mass function* is obtained, very similar to a probabilistic modeling especially when $|A| = 1$ ($|\cdot|$ is the cardinality) which withdraws the interest in using belief functions. An example is used in [25] for clustering purposes. It is evaluated according to a distance d_{ij} between an object j and a given cluster C_i with C being the discernment frame:

$$\begin{cases} m_j(\{C_i\}) = \alpha\varphi(d_{ij}), \\ m_j(C) = 1 - \alpha\varphi(d_{ij}), \\ m_j(A) = 0 \forall A \in 2^C \setminus \{C, \{\bar{C}\}\}. \end{cases} \quad (7)$$

$\varphi(\cdot)$ is a decreasing function based on d_{ij} such that the higher d_{ij} , the lower the confidence in the membership of the object to the class C_i . α is a source reliability-related factor. Rombaut [9] proposed an extension of Eq. (7) based on the notion of *specialized source* supporting only one hypothesis A of Θ :

$$\begin{cases} m_j(A) = \alpha\varphi(d_{ij}), \\ m_j(\bar{A}) = \alpha(1 - \varphi(d_{ij})), \\ m_j(\Theta) = 1 - \alpha. \end{cases} \quad (8)$$

As in Eq. (7), $\varphi(\cdot)$ is a decreasing function. Such *bba* can be self-conflicting since evidence can be assessed at a time on A and \bar{A} , meaning that both could be the solution. It has been widely used in MOA as it models imprecision and uncertainty. Similarly to Eq. (8), Gruyer *et al.* defined a *bba* which does not simultaneously express belief and disbelief in a hypothesis, i.e., a mass cannot

be generated at the same time on A and \bar{A} . Its main advantage is to avoid auto-conflict. More details can be found in [7].

The *simple support* model has the advantage of expressing the source's un-
230 certainty through the mass on the whole set which is not the case for the *comple-*
mentary mass function. It also necessitates a specific information as it indicates
the belief on a single proposition (singleton or set). In counterpart, none of these
bbas are appropriate when a source expresses belief on several propositions. The
complementary mass model obtained from Eq. (6) is able to deal with singletons
235 or composite propositions (sets). The main disadvantage of this model is that it
generates auto-conflict because of the simultaneous belief in a proposition and
its complement. Being very similar to a probabilistic modeling, these solutions
do not take advantage of the extensions provided by the evidence theory.

Specialized sources (cf. Eq. (8)) provide a more general framework since
240 they model the mass on a proposition (singleton or set), its complement and
ignorance. One major advantage is their adaptability since they can reproduce
optimistic, pessimistic or neutral behaviors thanks to the model parameters [7].
The counterpart is the parameter tuning or optimization required according
to the problem. When evidence conflicts, it might also raise a high conflict
245 because it comes down to combining two *simple support* functions. In this
context, the non-antagonist model from [7] offers a solution to the auto-conflict
as it independently describes a proposition and its complement on separate
intervals. Nevertheless, this specificity does not allow to optimally deal with
inaccuracy in the sensor data and uncertainty. The antagonist model (Eq. (8))
250 being less constraining, it fits better to the case of use described in this paper.

3.3. Refinement and Coarsening

Refinement and coarsening are important operations applied in order to ex-
press source's knowledge on the same discernment frame. Often, two sources can
express evidence on two distinct or complementary bodies of evidence. More-
255 over, any expression of a discernment frame Θ can generally be divided in differ-
ent propositions [4] since the discernment frame's granularity is usually chosen

by convention. In order to manage the pieces of knowledge and combine the information, it is necessary to have a larger or a more specific frame by the use of these operations.

260 Considering two frames Ω and Θ , a refinement $\rho(\cdot)$ of Θ to Ω is defined as a mapping $\rho(\cdot) : 2^\Theta \rightarrow 2^\Omega$ which verifies the following statements [4]:

$$\{\rho(\{\theta\}), \theta \in \Theta\} \subseteq 2^\Omega \text{ is a partition of } \Omega \quad (9)$$

$$\{\rho(A), A \subseteq \Theta\} = \bigcup_{\theta \in A} \{\rho(\{\theta\}), \theta \in \Theta\} \quad (10)$$

Therefore, Θ is a coarsening of Ω and Ω is a refinement of Θ . Considering a mass $m^\Theta(\cdot)$ expressing belief on Θ , the resulting mass $m^{\Theta\uparrow\Omega}(\cdot)$ elaborated by a refinement to Ω is found by applying a *vacuous extension* of $m^\Theta(\cdot)$ in Ω :

$$m^{\Theta\uparrow\Omega}(B) = \begin{cases} m^\Theta(A) & \text{if } B = \rho(A), A \subseteq \Theta, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Inversely, coarsening is denoted by the \downarrow operator as follows:

$$m^{\Theta\downarrow\Omega}(B) = \sum_{B \subseteq \Omega \setminus \bar{\rho}^{-1}(B)=A} m^\Omega(B), A \subseteq \Omega. \quad (12)$$

3.4. Combination of Evidence

This step is crucial for data fusion as it merges the various pieces of evidence delivered by the information sources. One of the main properties due to the evidential combination is the potential appearance of conflict. According to [15, 26], conflict can have several origins: aberrant measurements, inconsistent belief models and the multiplicity of sources. Moreover, conflict can raise from the non-reliability of sources and non-exhaustiveness of the discernment frame [27] and can be used as additional information in the fusion process [15].

270 The exhaustiveness of the discernment frame is made with the *close world* assumption implying $m(\emptyset) = 0$. In real case applications, it is rather hard or impossible to identify all solutions of a given problem. The frame of discernment exhaustiveness can thus not be guaranteed. Therefore, to treat the related

conflict, two additional frameworks have been introduced: the *open world* (*ow*)
 275 [6] and the *extended open world* (*eow*) [9, 10]. In the *ow*, sources are reliable
 and the conflict mass $m(\emptyset)$ represents the non-exhaustiveness of the discernment
 frame. In this context, the conflict is considered as the unknown hypotheses.
 This also applies to the TBM where the conflicting mass is not forced to zero
 after combination. In the *eow*, an additional hypothesis * is added to the dis-
 280 cernment frame in order to include all unknown hypotheses so that Θ is again
 exhaustive. The conflicting mass is then related either to the sources' unrelia-
 bility or the discordance between the data [15]. Therefore, the *eow* framework
 is adopted in this application.

Combining data when evidence conflicts has been largely investigated and
 285 several combination rules have emerged. This section briefly recalls those which
 have been used in this work and evaluated in Section 5. For further details,
 the reader can refer to [15, 28]. The chosen rules are all based on the conjunc-
 tive operator. Disjunctive rules are disregarded because they induce a loss of
 specificity and can only be used when at least one source is reliable.

Under the assumption of exhaustiveness and exclusivity of Θ , the fundamen-
 tal combination rule is Dempster's rule [4] or orthogonal sum (OS). It requires
 independent and reliable sources and is formulated as follows:

$$m_{\oplus}(A) = \frac{1}{1 - \kappa} \sum_{A_1 \cap \dots \cap A_p = A} \prod_{j=1}^p m_j(A_j), \quad (13)$$

where κ is the conflict such that:

$$\kappa = m(\emptyset) = \sum_{A_1 \cap \dots \cap A_p = \emptyset} \prod_{j=1}^p m_j(A_j). \quad (14)$$

290 Proposed by Smets [6] in the TBM in order to preserve the conflict, the
 conjunctive combination does not require the exhaustiveness of the discernment
 frame. Mathematically, it is an unnormalized orthogonal sum:

$$\begin{cases} m_{\cap}(A) = \sum_{A_1 \cap \dots \cap A_p = A} \prod_{j=1}^p m_j(A_j), \\ m_{\cap}(\emptyset) = \kappa = \sum_{A_1 \cap \dots \cap A_p = \emptyset} \prod_{j=1}^p m_j(A_j). \end{cases} \quad (15)$$

Conceptually, when all sources are reliable, the belief on the empty set $m_{\cap}(\emptyset)$ informs about a solution outside of Θ . It is useful to precise that Eq. (13) and Eq. (15) describe associative and commutative operators. With Yager's rule (Y) [29], the conflicting mass $m_{\cap}(\emptyset)$ obtained after a conjunctive combination is assigned to Θ . Therefore, when evidence conflicts, we are ignorant about the problem solution:

$$\begin{cases} m_Y(A) = m_{\cap}(A), & \forall A \subset \Theta, \\ m_Y(\Theta) = m_{\cap}(\Theta) + m_{\cap}(\emptyset). \end{cases} \quad (16)$$

This operator is associative but not commutative requiring a combination sequence to be defined *a priori*.

Also known as the mixed rule, Dubois and Prade's rule (DP) [30] merges the conjunctive and disjunctive operators. The conflicting mass raised by $A \cap B = \emptyset$ is attributed to the proposition $A \cup B$. For two pieces of evidence $m_1(\cdot)$ and $m_2(\cdot)$, the combined mass $m_{DP}(\cdot)$ is:

$$m_{DP}(X) = \sum_{A \cap B = X} m_1(A)m_2(B) + \sum_{\substack{A \cup B = X \\ A \cap B = \emptyset}} m_1(A)m_2(B). \quad (17)$$

295 This operator is not commutative thus when fusing several *bbas*, it is important to define the fusion sequence.

The Proportional Conflict Redistribution (PCR) rules have been proposed by Dezert and Smarandache [31]. These operators allow a proportional redistribution of the partial conflict according to the concerned focal elements. Several extensions have been introduced by Martin and Osswald to increase the number of experts, such as the PCR6 [27]:

$$m_{\text{PCR6}}(X) = m_{\cap}(X) + \sum_{\substack{\bigcap_{k=1}^{M-1} Y_{\sigma_i(k)} X = \emptyset \\ (Y_{\sigma_i(1)}, \dots, Y_{\sigma_i(M-1)}) \in (2^{\Theta})^{M-1}}} \left(\frac{\prod_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})}{m_i(X) + \sum_{j=1}^{M-1} m_{\sigma_i(j)}(Y_{\sigma_i(j)})} \right). \quad (18)$$

where $Y_j \in 2^{\Theta}$ is the response of the expert j , $m_j(Y_j)$ the associated belief

function and σ_i counts from 1 to M :

$$\begin{cases} \sigma_i(j) = j & \text{if } j < i, \\ \sigma_i(j) = j + 1 & \text{if } j \geq i. \end{cases} \quad (19)$$

3.5. Decision-making

The last step of the fusion scheme is decision-making, i.e. finding the best solutions under uncertainty among the potential hypotheses in Θ . The *bbas* combination leads to evidences placed either on singletons or on unions. The decision performs a mapping from 2^Θ to Θ in order to retain the best proposition of Θ related to the problem of concern. In order to handle uncertainty, this process is based on the optimization of an evidential function.

Classical decision criteria are based on maximizing the credibility function which is considered to be pessimistic or the plausibility function which is optimistic. A common approach adopted by Smets [32] consists in the transformation of evidence to probabilities. It provides an intermediate solution between the maximum of credibility and plausibility. The pignistic probability is computed through the relation:

$$BetP(H_j) = \sum_{\substack{A \in 2^\Theta \\ H_j \subseteq A}} \frac{m^\Theta(A)}{|A|(1 - m^\Theta(\emptyset))}, \quad (20)$$

with $m^\Theta(\emptyset) < 1$ and $|A|$ the cardinality of A . Therefore, the decision by maximizing $BetP(\cdot)$ according to the problem's constraints is given by:

$$H_{BetP} = arg \max_{H_j \in \Theta} BetP(H_j). \quad (21)$$

3.6. Representation of Evidence in MOA

This section recalls fundamental elements in object association using belief functions and already described by some of the authors [14, 24].

Considering the target and track sets introduced in Section 2.2 (see (1) and (2)) Θ_{X_i} and Θ_{Y_j} respectively represent two discernment frames gathering the M tracks, N targets, appearance and disappearance propositions are defined.

Belief degrees are set on the association potentiality of each pair of target and track. A pairwise association concerns the evaluation if “yes” or “no” these objects are related to each other. For a given source, its belief will then be expressed on these two propositions defining the frame of discernment $\Theta_{ij} = \{yes, no\}$ which will be simplified to $\Theta_{ij} = \{y, n\}$. Three pieces of evidence are constructed: the belief in the association $m_{i,j}(\{y\})$, in the non association $m_{i,j}(\{n\})$ and $m_{i,j}(\{\Theta_{ij}\})$ stands for the ignorance. Obviously, both frames Θ_{X_i} and Θ_{Y_j} are refinements of Θ_{ij} (see Fig. 3). By applying a vacuous extension

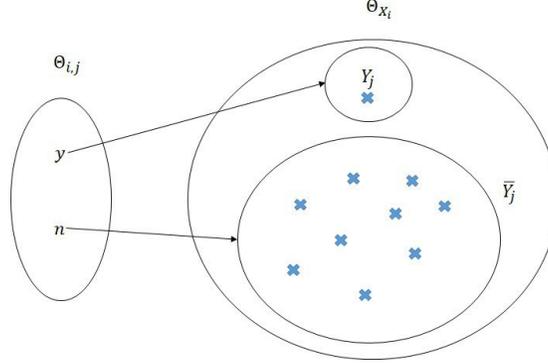


Figure 3: Refinement enabling the transfer from Θ_{ij} to Θ_{X_i} [12].

(cf. Section 3.3), the belief can be expressed on a common frame:

$$\begin{cases} m_j^{\Theta_{X_i}}(Y_j) = m_{i,j}(\{y\}), \\ m_j^{\Theta_{X_i}}(\bar{Y}_j) = m_{i,j}(\{n\}), \\ m_j^{\Theta_{X_i}}(\Theta_{X_i}) = m_{i,j}(\Theta_{ij}). \end{cases} \quad (22)$$

310 It is worth noting that \bar{Y}_j is the non association. It corresponds to the set containing all tracks except Y_j , i.e. $\Theta_{X_i} \setminus Y_j = \{Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_M, Y_*\}$. An object pairing is assessed based on two assumptions:

- Only one target is associated to a track (*one-to-one* association),
- There might be appearances and disappearances of tracks.

315 To attribute masses to respective sources using Eq. (22), a proper modeling needs to be initiated depending on the nature and characteristics of the data.

The choice of models according to those described in Section 3.2 will be discussed in Section 4.

The main criteria for evidential decision-making is the pignistic probability $BetP(\cdot)$ (cf. Section 3.5). This measure is mostly used because it converts mass functions to probabilities convenient for selecting the most eligible hypothesis. When using a conjunctive combination (cf. Section 3.4), the combined masses can be obtained from the initial *bbas* as recalled in [14]. Afterwards, based on Eq. (20), the pignistic probability for each object X_i is given by:

$$BetP_{X_i}(Y_j) = m_j^{\Theta_{X_i}}(Y_j) \prod_{\substack{a=1 \\ a \neq j}}^M (1 - m_a^{\Theta_{X_i}}(Y_a)) + \sum_{\substack{A \subseteq \Theta_{X_i} \\ Y_j \in A \\ |A| > 1}} \frac{1}{|A|} \prod_{\substack{b=1 \\ Y_b \in A}}^M (m_b^{\Theta_{X_i}}(\Theta_{X_i})) \prod_{\substack{b=1 \\ Y_b \notin A}}^M (m_b^{\Theta_{X_i}}(\bar{Y}_b)), \quad (23)$$

$$BetP_{X_i}(Y_*) = \prod_{a=1}^M (m_a^{\Theta_{X_i}}(\bar{Y}_a)) + \sum_{\substack{A \subseteq \Theta_{X_i} \\ Y_* \in A \\ |A| > 1}} \frac{1}{|A|} \prod_{\substack{b=1 \\ Y_b \in A}}^M (m_b^{\Theta_{X_i}}(\Theta_{X_i})) \prod_{\substack{b=1 \\ Y_b \notin A}}^M (m_b^{\Theta_{X_i}}(\bar{Y}_b)). \quad (24)$$

The conflicting mass function in Eq. (20) is defined by:

$$m_j^{\Theta_{X_i}}(\emptyset) = 1 - \left[\prod_{a=1}^M \xi_a + \sum_{a=1}^M m_a^{\Theta_{X_i}}(Y_a) \prod_{\substack{b=1 \\ b \neq a}}^M \eta_b \right] \quad (25)$$

with $\xi_a = (1 - m_a^{\Theta_{X_i}}(Y_a))$ and $\eta_b = (1 - m_b^{\Theta_{X_i}}(Y_b))$. The resulting pignistic probabilities are therefore normalized by $(1 - m_j^{\Theta_{X_i}}(\emptyset))$. It can be especially noticed that a decision of appearance/disappearance is related to the belief in the non-association combined with the ignorance (see Eq. (24)). Since the association is bi-directional ($X_i \Rightarrow Y_j$ and $Y_j \Rightarrow X_i$), the pignistic transformation is realized for both cases which results in two probability matrices $BetP_{X_i}(\cdot)$ and $BetP_{Y_j}(\cdot)$ shown in Table 1 and Table 2. Each line defines the probabilities of the associations of X_i with $Y_1 \dots Y_M, *$ and *vice versa*.

$BetP_{X_i}(\cdot)$	Y_1	\dots	Y_M	$*$
X_1	$BetP_{X_1}(Y_1)$	\dots	$BetP_{X_1}(Y_M)$	$BetP_{X_1}(*)$
X_2	$BetP_{X_2}(Y_1)$	\dots	$BetP_{X_2}(Y_M)$	$BetP_{X_2}(*)$
\vdots	\vdots	\vdots	\vdots	\vdots
X_N	$BetP_{X_N}(Y_1)$	\dots	$BetP_{X_N}(Y_M)$	$BetP_{X_N}(*)$

Table 1: Target-to-Track pignistic matrix

$BetP_{Y_j}(\cdot)$	X_1	\dots	X_N	$*$
Y_1	$BetP_{Y_1}(X_1)$	\dots	$BetP_{Y_1}(X_N)$	$BetP_{Y_1}(*)$
Y_2	$BetP_{Y_2}(X_1)$	\dots	$BetP_{Y_2}(X_N)$	$BetP_{Y_2}(*)$
\vdots	\vdots	\vdots	\vdots	\vdots
Y_M	$BetP_{Y_M}(X_1)$	\dots	$BetP_{Y_M}(X_N)$	$BetP_{Y_M}(*)$

Table 2: Track-to-Target pignistic matrix.

4. Multiple Feature Fusion

This section describes the multiple feature hierarchical fusion. The aim is to allow different information sources to distinctively describe dynamic objects in a complex driving environment. Firstly, it highlights how these data are combined in order to obtain a global view of the potential pairings. In a second step, we show how the belief masses of the heterogeneous object data are obtained thanks to the *bba* introduced previously.

4.1. Fusion Structure

The need for multiple sensors is recurrent in a wide number of applications. This necessitates an efficient fusion considering the imprecision and the incompleteness of the respective data. In this study, the focus will primarily be laid on the inconsistency of reliable sources, i.e., when their belief is based on information which might not always be conclusive at decision-making.

The hierarchical multi-feature fusion is suggested as a solution to ambiguity resolution in MOA. The aim is to gather enough knowledge about the objects

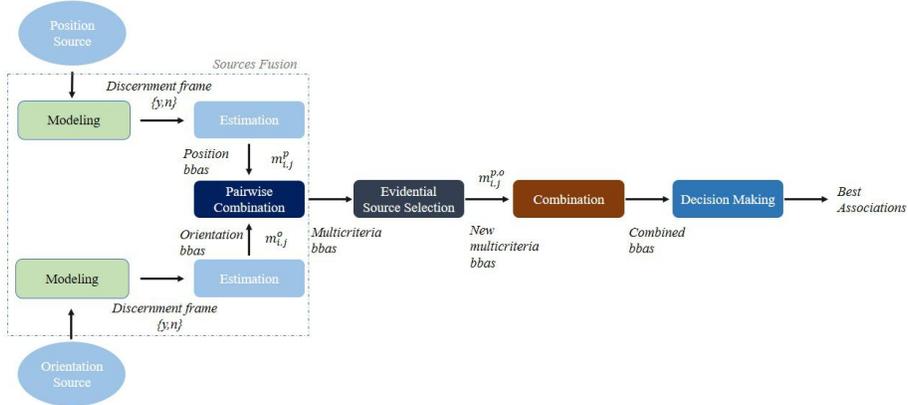


Figure 4: Hierarchical multiple feature fusion architecture.

behavior and displacement in the scene, i.e. their position/scale and direction of motion. The fusion strategy is applied according to Fig. 4. First, both similarity measures describing the object position and its direction of motion are computed separately. These measures are then used to obtain the respective *position* $m_{i,j}^p$ and *orientation* $m_{i,j}^o$ bbas. This section details each step and Algorithm 1 gives an overview of the process.

Since both sources provide distinct and complementary information about the pairwise association of X_i and Y_j , it is convenient to fuse them (Fig. 4 Pairwise Combination step). For a pair of objects X_i and Y_j , this first fusion level aims at gathering the partial information given by the position and orientation source in order to obtain a mass function $m_{\odot_{i,j}}^{p,o}(\cdot)$ describing the total belief in the pairwise association:

$$m_{\odot_{i,j}}^{p,o}(\cdot) = m_{i,j}^p(\cdot) \odot m_{i,j}^o(\cdot), \quad (26)$$

with $i \in I = \{1 \dots N\}$, $j \in J = \{1 \dots M\}$ and \odot can refer to any combination operator shown in Section 3.4. This step is sequentially performed among all association pairings.

Whatever the association method or framework employed, it is always suitable to limit/reduce the number of assignment hypotheses. The ‘‘Evidential Source Selection’’ in Fig. 4 avoids unnecessary association evaluations and thus

limits the computation requirements. Gating is a standard method used in MTT
 355 to identify the neighboring targets for a pairing with a particular track. A similar strategy was introduced by the authors in [24] to implement an evidential selection. The objective is to rely on the source data to discard those having a low belief in a pairing hypothesis and only maintain the most pertinent ones. More details are given in [24] in which the experimental validation shows the
 360 reduction in computation cost coupled to high association performances.

A second level of combination is performed (cf. Fig. 4) in order to allocate a mass $m_{\odot_{i,.}}^{p,o}(\cdot)$ or $m_{\odot_{.,j}}^{p,o}(\cdot)$ to express the belief applied to the association of resp. a target i /track j to all resp. tracks/targets:

$$\begin{cases} m_{\odot_{i,.}}^{p,o}(\cdot) = m_{\odot_{i,1}}^{p,o}(\cdot) \odot \dots \odot m_{\odot_{i,M}}^{p,o}(\cdot), \\ m_{\odot_{.,j}}^{p,o}(\cdot) = m_{\odot_{1,j}}^{p,o}(\cdot) \odot \dots \odot m_{\odot_{N,j}}^{p,o}(\cdot), \end{cases} \quad (27)$$

where N and M are the number of targets and tracks respectively. This combination is sequentially applied over all targets ($i \in \{1 \dots N\}$) and tracks ($j \in \{1 \dots M\}$). Finally, the decision making is performed thanks to the transformation of the combined belief masses $m_{\odot_{i,.}}^{p,o}(\cdot)$ and $m_{\odot_{.,j}}^{p,o}(\cdot)$ into pignistic probabilities with Eq. (23) and (24). The best hypotheses are found based on the
 365 optimization presented in Section 3.5.

4.2. BBA from Sensor Data: Object Position Modeling

In this section, a recall of [24] on how the relative position between objects can be treated to build pairwise association *bbas* is presented. For the position, the *bbas* are defined according to Eq. (8). The position attribute is, in this application, the most informative w.r.t. the association objective. By evaluating the distance between two objects in a common spatial frame, it will be possible to define if they are associated or not or if there is a doubt. Roughly speaking, this model extends the fact that an object cannot be at two different locations in the scene and two different objects cannot have exactly the same position at a given time. The masses are based on statistical distances (Euclidean, Mahalanobis,

Algorithm 1 Multiple Feature Fusion

Require: $d_{ij}, \Delta\psi_{ij}, i \in I = \{1, \dots, N\}, j \in J = \{1, \dots, M\}$.

Ensure: Hierarchical and sequential combination to compute $m_{\odot_{i..}}^{p,o}(\cdot)$ and

$m_{\odot_{..j}}^{p,o}(\cdot)$.
 {Pairwise Combination}

for $i = 1$ to N **do**

for $j = 1$ to M **do**

$$m_{\odot_{i,j}}^{p,o}(\cdot) = m_{i,j}^p(\cdot) \odot m_{i,j}^o(\cdot)$$

end for

end for

{Evidential Source Selection (see [24])}

{Sequential Combination}

for $i = 1$ to N **do**

$$m_{\odot_{i..}}^{p,o}(\cdot) = m_{\odot_{i,1}}^{p,o}(\cdot) \odot \dots \odot m_{\odot_{i,M}}^{p,o}(\cdot)$$

end for

for $j = 1$ to M **do**

$$m_{\odot_{..j}}^{p,o}(\cdot) = m_{\odot_{1,j}}^{p,o}(\cdot) \odot \dots \odot m_{\odot_{N,j}}^{p,o}(\cdot)$$

end for

etc.) between each pair of objects [25]:

$$\begin{cases} m_{i,j}^p(\{y\}) = \alpha' \exp^{-\gamma' d_{ij}^{\beta'}}, \\ m_{i,j}^p(\{n\}) = \alpha' (1 - \exp^{-\gamma' d_{ij}^{\beta'}}), \\ m_{i,j}^p(\Theta_{i,j}) = 1 - \alpha', \end{cases} \quad (28)$$

where $0 < \alpha' < 1$ is the sensor *a priori* known reliability. $\gamma' \in \mathbb{R}^*$ and $\beta' \in \mathbb{N}^*$ are weighting parameters. Generally speaking, the model parameters are defined heuristically [25]. In Section 5.2, the identification of these coefficients will be discussed. In this application, each object is defined by a bounding box in the image coordinates as shown in Fig. 5. The position and scale of an object can be evaluated based on the top left corner and the bottom right corner of the bounding box. Therefore, the similarity measure d_{ij} is defined as follows:

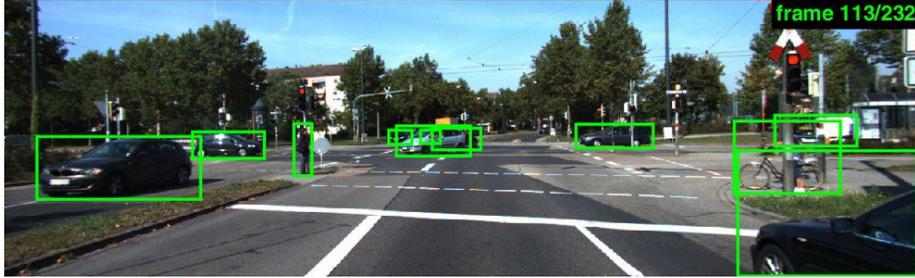


Figure 5: Example of various object detections (Extracted from KITTI Seq.02 [22]).

$$d_{ij} = \frac{d_{ij}^{tl} + d_{ij}^{br}}{2}, \quad (29)$$

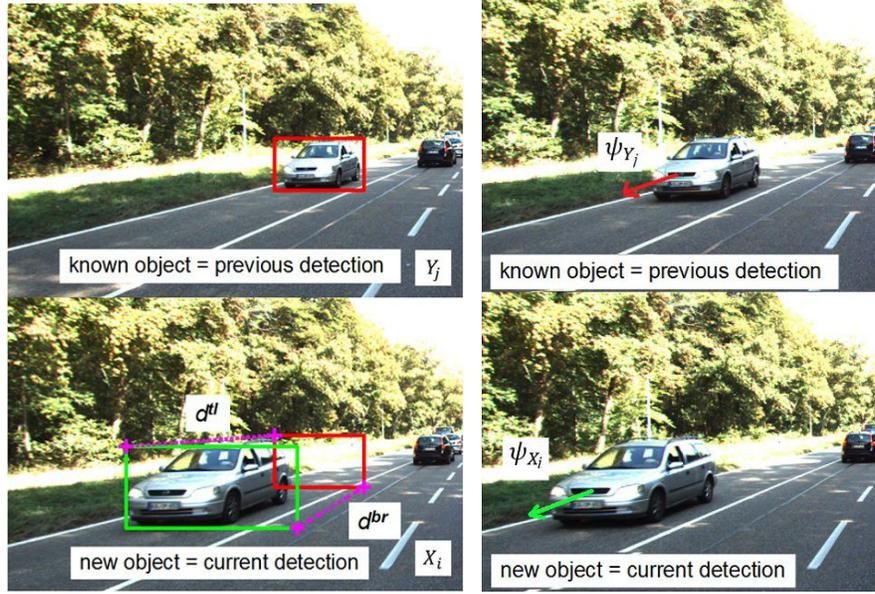
375 where d_{ij}^{tl} and d_{ij}^{br} stand for the Euclidean distances in the sensor frame between the top left (*tl*) and bottom right (*br*) corners of the X_i and Y_j bounding boxes respectively as shown in Fig. 6a.

4.3. BBA from Sensor Data: Object Motion Direction Modeling

Assuming that the direction of motion of each track Y_j and target X_i is
 380 available throughout time by a given sensor (here a LiDAR), we propose to compute the similarity measure $\Delta\psi_{ij}$ as follows:

$$\Delta\psi_{ij} = |\psi_{X_i} - \psi_{Y_j}|, \quad (30)$$

where $\Delta\psi_{ij}$ is the relative direction of motion of objects X_i and Y_j while ψ_{X_i} and ψ_{Y_j} are the direction of motion of each of them (Fig. 6b). This measure is therefore used to generate the adequate *bba*. In a similar way than for the position *bba*, the representation of a *specialized source* model is retained. However, it is obvious that for object pairings, the orientation is less pertinent or informative than the position. In fact, two different objects can move in the same direction leading to a small relative orientation. In this case, these objects should not be associated. A first idea would be to define a *bba* supporting either the non association or the ignorance as performed in [33]. This model can be



(a) Position and size attribute

(b) Direction of motion attribute

Figure 6: Attribute determination (Extracted from KITTI Seq. 18 [22]).

built from the antagonist mass function Eq. (8):

$$\text{Model 1: } \begin{cases} m_{i,j}^o(\{y\}) = 0, \\ m_{i,j}^o(\{n\}) = \alpha(1 - \exp^{-\gamma\Delta\psi_{ij}^\beta}), \\ m_{i,j}^o(\Theta_{ij}) = 1 - \alpha(1 - \exp^{-\gamma\Delta\psi_{ij}^\beta}). \end{cases} \quad (31)$$

Eq. (31) supports the non-association when object have different direction in the sensor frame or at least, it is ignorant. Its drawback is that it will never be confident in the association of objects even if they are matching. This might alter the decision-making process and select false associations (an example is presented in Section 5). To cope with this inconvenience, the direction of motion source can be expressed defined by Eq. (32) similarly to the position *bba*. One objective of this paper is to study both models on different real scenarios in

order to evaluate their impact on the association process (see Section 5).

$$\text{Model 2: } \begin{cases} m_{i,j}^o(\{y\}) = \alpha \exp^{-\gamma \Delta \psi_{ij}^\beta}, \\ m_{i,j}^o(\{n\}) = \alpha(1 - \exp^{-\gamma \Delta \psi_{ij}^\beta}), \\ m_{i,j}^o(\Theta_{i,j}) = 1 - \alpha. \end{cases} \quad (32)$$

5. Experimental Results

This section evaluates the proposed multi-feature fusion on real data from KITTI [22]. It largely extends [33] in which the authors rated the orientation
 385 model 1 from Eq. (31) on a pedestrian detection and tracking sequence from KITTI. Here, tests are conducted with both orientation *bbas* on multiple scenarios (urban and highway traffic, etc.). First, the dataset used for validation and the assessment criteria are presented. In a second step, the *bbas* parameter identification is discussed and the position and orientation *bbas* validity is
 390 shown. Finally, a quantitative evaluation of the fusion framework and a discussion related to conflict management are proposed.

5.1. Dataset Description and Evaluation Criteria

KITTI provides a set of sequences containing labeled detections of pedestrians, cars, cyclists, vans, and trucks in various road scenes (highway, urban
 395 driving, etc.) as ground truth. Data from several sensors (inertial measurement units, GPS, cameras, LiDAR, etc.) mounted on a manually driven car are available. The results described here focus on the image and LiDAR data. The detections are defined by object tracklets which contain the dimensions of the 2D bounding boxes, occlusion states on the image frame, etc. For each detection,
 400 a 3D bounding box defines the set of LiDAR data that coincides with the object. The orientation angles used here correspond to the motion direction of these 3D detections. Several sequences are used due to their object and road context heterogeneity. Table 3 gives the sequence details and Fig. 7 a screenshot of the driving situations. They vary according to the number and nature
 405 of detections as well as the vehicle’s speed during data recording. For instance,

Sequence 17 contains a set of frames which was captured when the car was stationary. The acquisition from Sequence 08 is majorly established between 30 to 60 *km/h* and more. In Sequence 18, the vehicle’s speed was mostly less than 30 *km/h*. A total number of nearly 2 000 associations are evaluated for these 3 sequences.

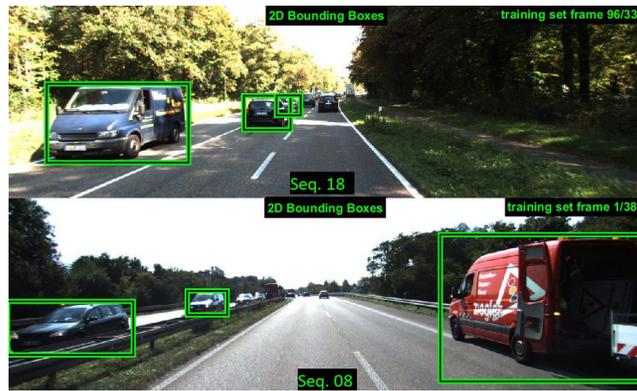


Figure 7: Example frames from Sequence 08 (bottom) and 18 (top).

	Seq.08	Seq.17	Seq.18
Number of frames	390	145	339
Number of associations	492	434	1130
Max vehicle speed (km/h)	62	0	55
Min vehicle speed (km/h)	38	0	0
Speed < 30 km/h (%)	0	100	66
30 < Speed < 60 km/h (%)	86	0	34
Speed > 60 km/h (%)	14	0	0

Table 3: KITTI Image sequence characteristics.

	Pedestrian	Car	Cyclist	Van
Seq.08	✓	✓	✓	✓
Seq.17	✓	-	-	-
Seq.18	✓	✓	✓	✓

Table 4: Object class in the KITTI sequences.

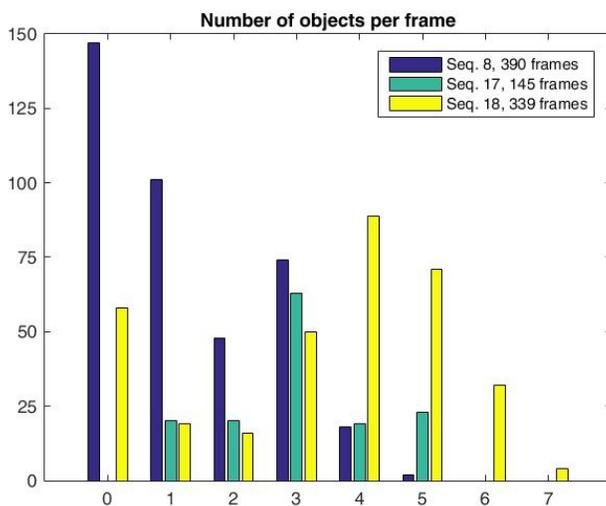


Figure 8: Number of detected objects per frame w.r.t the used sequences.

Fig. 8 shows the number of detected objects per frame according to the used sequences. For instance, Seq.08 and Seq.17 contain at most 5 objects. However, sequence 18 contains frames having up to 7 objects and most of the frames contain 4 to 5 objects. Therefore, Seq.18 has the most associations to process, 1130 against less than 500 for Seq.08 and Seq.17. Finally, Table 4 describes the nature/class of objects labeled in the respective sequences. Their heterogeneity can be easily attested.

The evaluation of the approach is established according to the rate of true associations known as the recall. The recall is defined as the fraction of matched pairs that are correct and the fraction of true object pairs that were matched [13]:

$$Recall(\%) = \frac{|\{true\ pairs\} \cap \{matched\ pairs\}|}{\{matched\ pairs\}}. \quad (33)$$

5.2. Model Identification

Considering the importance of parameters α'/α , β'/β and γ'/γ in the *bbas* given by Eq. (28), Eq. (31) and Eq. (32), the determination of these values according to the nature of the similarity measure is investigated. Basically, each parameter has a specific meaning. α'/α translate the sensor's reliability, β'/β and γ'/γ are weighting parameters function of the targeted application. α'/α depends on the source providing the data to be gathered, here the sensors. In this case, due to the high quality of these sensors, they are considered reliable. The parameters are set to 0.9 ($\alpha = \alpha' = 0.9$), standard values for this kind of application. This is a good compromise since it allows to keep some ignorance even if the source is confident in the association (i.e. when d_{ij} is close to 0). β'/β has very little impact on the performance of the association, it can be fixed either to 1 or 2 [25]. Here, it is decided, for comparison purposes, to follow the standard choice for MOA which is to set $\beta = \beta' = 1$ [13].

γ'/γ is very important because it represents the decreasing rate of the association confidence. It should be chosen considering the nature of the similarity measure and its variation range. Here, a sensitivity study is made to determine these factors according to the recall (Eq. (33)). The analysis is devoted to the object position modeling but the same is applied to identify the object motion direction model. Figure 9 shows the variation of the recall according to the KITTI Sequence 08, 17 and 18. It can be seen that the most significant results are obtained for $\gamma' = 0.01$. This value is not only convenient for the sequences that have been tested but it can be generalized to any sequence w.r.t. the resolution of the considered images. Since the similarity measure d_{ij} corresponds to the relative distance of the recorded positions in pixels, its values can vary at most in the whole frame size, i.e. $d_{ij} \in [0...1294.7]$. An example of the variation of d_{ij} according to each frame of Sequence 08 is illustrated in Figure 10. The confidence in the association rises when this measure is low and tends to zero

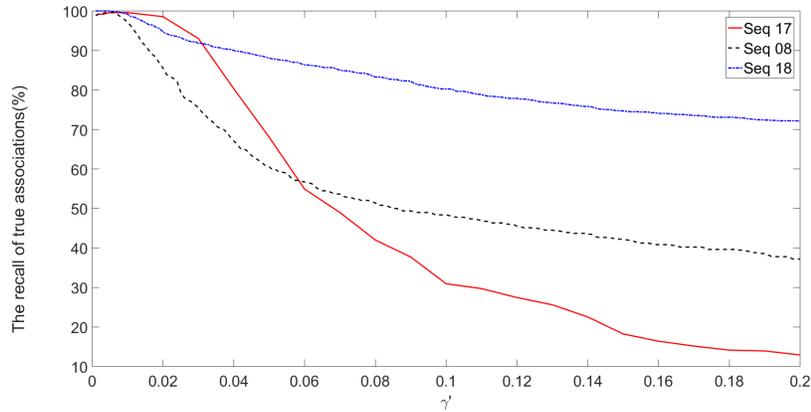


Figure 9: Influence of the position decreasing rate γ' on the association performance.

when d_{ij} attains its maximum value. The factor γ' monitors the variation of the confidence w.r.t. the similarity as shown in Figure 11.

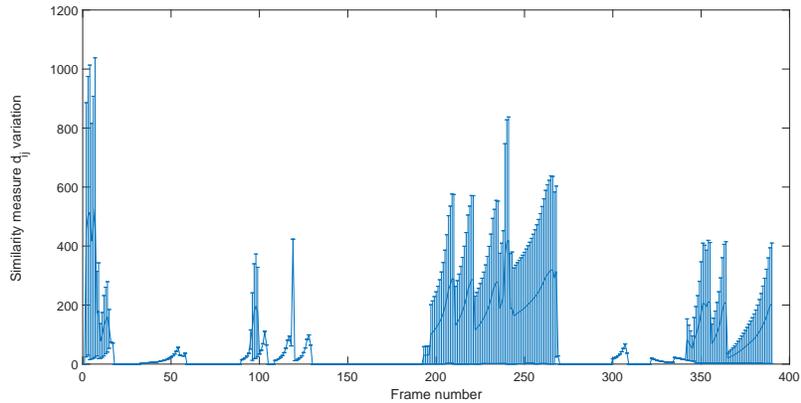


Figure 10: Minimum and maximum values of d_{ij} for each frame from Seq.08.

To measure the impact of γ' on the mass model, Figure 11 shows three different configurations. It can be seen obviously that the larger γ' is, the faster the association mass decreases with respect to the similarity measure d_{ij} . It can be noticed that for $\gamma' = 0.02$ and $\gamma' = 0.1$, the association masses are quickly tending to 0 which explains the decreasing recall in Figure 9. This configurations would lead in a loss of confidence in the association with small variations of the

similarity measure and would not be appropriate in our case.

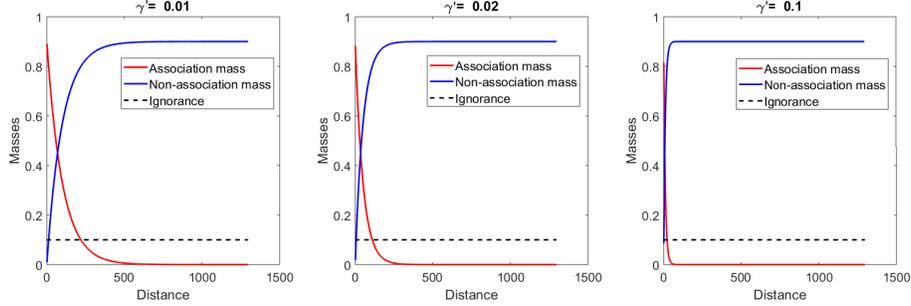


Figure 11: Influence of γ' on belief assignment.

455 By applying the same approach to the orientation source model, the parameters α, β and γ have been fixed to resp. 0.9, 1.0 and 1.5.

5.3. Orientation Motion Model Validation

The orientation source is evaluated according to both mass models presented in Section 4.3 and the parameters α, β and γ previously obtained. This validation is performed in iso-conditions w.r.t the model parameters. This allows the source to distinguish slight variations of the object relative direction of motion. For this experiment, **Sequence 17** is used. In this scenario, five moving pedestrians are detected (cf. Fig. 12). Four of them (X_1 to X_4) are going in the same direction whereas X_5 , partially occluded by X_1 and X_2 , is walking in the opposite direction. Moreover, the objects have very close positions in the image frame. Tracks Y_1, Y_2, Y_3, Y_4 and Y_5 (occluded) should be associated to X_1, X_2, X_3, X_4 and X_5 respectively.

Fig. 13 presents the *bbas* generated with both orientation models corresponding to the belief supporting the relation between target X_1 (new object) to track Y_5 (known object). The belief assignment is made according to the relative direction of motion of objects detected between two successive frames as shown in Fig. 12. Model 2 distributes the knowledge on association, non-association, and ignorance (cf. Fig. 13) but is not confident in the association ($m_{1,5}^o(\{y\}) = 0.01$) of X_1 and Y_5 as they move in opposite directions. Moreover, both models are

475 equally confident in the non-association of X_1 to Y_5 . Finally, their amount of
 ignorance is very similar in this situation and both correctly describe the behav-
 ior of objects in the scene where X_1 cannot be associated to Y_5 when referring
 to their motion.

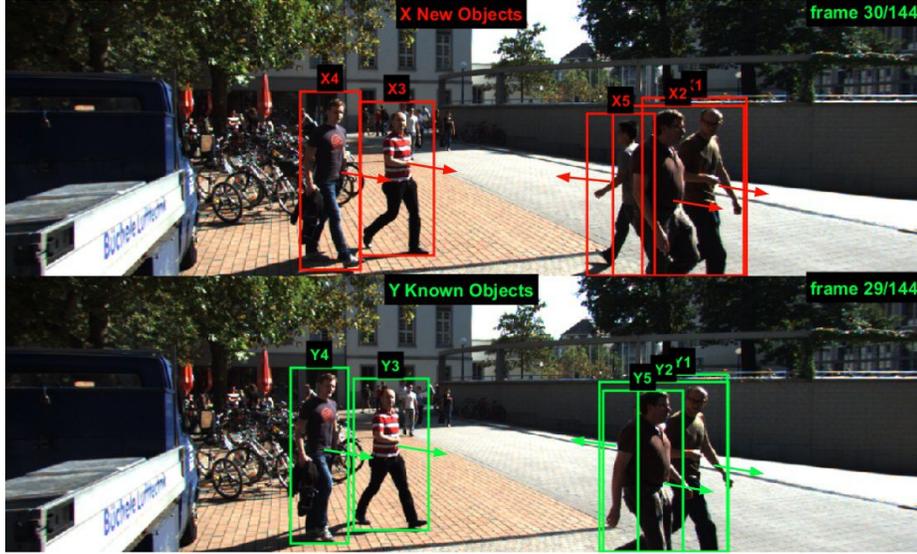


Figure 12: Pedestrian position and direction of motion: from known to new detections (lower to upper frame).

Consider now X_1 and Y_1 , a target and a track that should be associated
 480 (ground truth). Fig. 14 presents the masses $m_{1,1}^o(\cdot)$ issued with Model 1 and
 2. Model 2 is very confident about their association whereas Model 1 remains
 largely ignorant due to the close direction of motion: the mass on the whole set
 is close to the vacuous mass function. In this case, Model 2 is more informative
 about the association objective.

485 In order to evaluate the presented models for the MOA, they are both in-
 cluded in the hierarchical fusion described earlier. The process is evaluated
 on frame 30 (Fig. 12). First, the object position and size source provides the
 following associations (target-to-track | track-to-target):

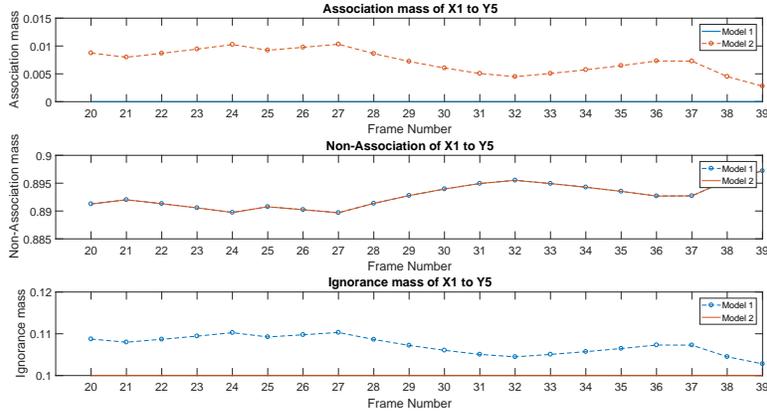


Figure 13: $m_{1,5}^o(\cdot)$ according to the orientation models.

X_1	X_2	X_3	X_4	X_5	Y_1	Y_2	Y_3	Y_4	Y_5
↓	⇓	↓	↓	⇓	↓	⇓	↓	↓	⇓
Y_1	Y_5	Y_3	Y_4	Y_2	X_1	X_5	X_3	X_4	X_2

Table 5: Association decisions with the position/size source on KITTI Seq. 17.

It can be seen that an ambiguity (represented by a double arrow) occurs in the association of $X_2 \rightarrow Y_5$ and $X_5 \rightarrow Y_2$ in both ways due to the close position and occlusion of these objects in the scene. The standalone orientation source generated according to Model 1 believes in the following assignments:

X_1	X_2	X_3	X_4	X_5	Y_1	Y_2	Y_3	Y_4	Y_5
⇓	↓	⇓	↓	⇓	⇓	↓	⇓	⇓	⇓
Y_*	Y_2	Y_*	Y_4	Y_*	X_*	X_2	X_*	X_*	X_*

Table 6: Association decisions with Model 1 on KITTI Seq. 17.

It is noticeable that many wrong decisions are made since this *bbm* mainly supports appearances (Y_*) and disappearances (X_*) of objects as shown by Eq. (24). It can be concluded that the ignorance given by a source supports mainly the appearance/disappearance probability, leading to incorrect associations. On the other hand, the standalone orientation source generated according

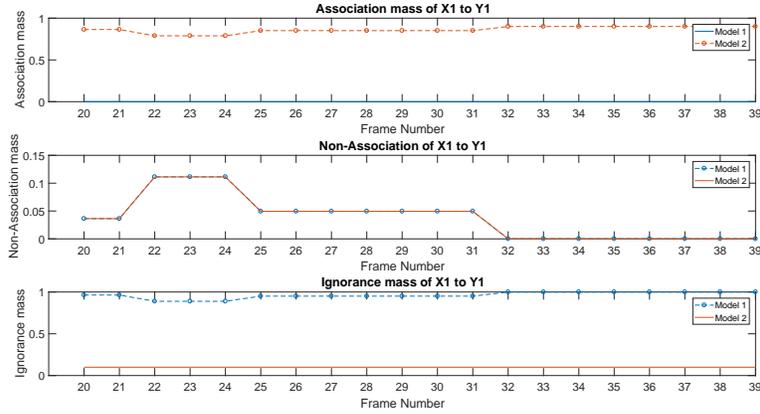


Figure 14: $m_{1,1}^o(\cdot)$ according to orientation models.

to Model 2 gives the following pairings:

X_1	X_2	X_3	X_4	X_5	Y_1	Y_2	Y_3	Y_4	Y_5
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Y_1	Y_2	Y_3	Y_4	Y_5	X_1	X_2	X_3	X_4	X_5

Table 7: Association decisions with Model 2 on KITTI Seq. 17.

All associations are correct in both ways. Even though some distinct objects
500 move in the same direction, for instance target X_3 and track Y_4 , the model
parameters allow the source to easily differentiate them. Finally, it will be
shown that better pairing results are obtained when using Model 2 to represent
the orientation information in the fusion process.

5.4. Performance Rating

Table 8 compares the performance of the standalone sources and the proposed
505 multi-feature fusion approach. The evaluation is done according to se-
quences 08, 17, and 18. To evaluate more effectively the proposed approach, the
estimation and combination steps are investigated. The results describe both
orientation *bbms* and two combination rules are investigated: the conjunctive
510 rule and the orthogonal sum (OS).

The results show that the orientation source attains significantly different rates of true associations with both models. In fact, for all sequences, Model 2 gives better performance than Model 1 as expected in Section 5.3. On the other hand, the position source scores a high rate of true associations for all sequences confirming that this is the most informative data for heterogeneous object pairing. Nevertheless, the complexity and variety of the scenes and of the related objects already leads to some wrong associations. An error-free pairing only with the position is not achievable mainly due to the occlusion cases or crossing scenarios. The best scores (100%) are obtained by the fusion strategy with Model 2 for the orientation source. However, an important factor to be considered is the combination operator. For instance, in **Sequence 18**, the enhancement of the recall is remarkable when normalizing the conflict in the two levels of combination. It raises from approximately 57% to 100%. Here, it can be explained by the absorbing property of the conflict during successive conjunctive combinations lowering the belief in the pairwise associations. This reveals the need for an investigation on the conflict management method and the comparison of different operators which will be exposed in the next section.

	Combination operator	Position	Orientation (Model 1)	<i>Fusion</i>	Orientation (Model 2)	<i>Fusion</i>
Seq. 08	Conjunctive	97.25	75.02	<i>97.05</i>	96.75	99.59
	OS	97.26	75.02	<i>97.05</i>	96.54	99.69
Seq. 17	Conjunctive	98.40	55.61	<i>99.09</i>	94.06	<i>90.18</i>
	OS	99.54	55.61	100	95.32	100
Seq. 18	Conjunctive	98.76	90.03	<i>98.58</i>	68.36	<i>57.29</i>
	OS	99.20	90.03	<i>98.62</i>	96.50	100

Table 8: Association scores by standalone sources and their combination.

5.5. Conflict Management

Considering the variety of combination rules in the evidence theory, it is necessary to guarantee the conformity of a chosen rule to the addressed problem. In the presented approach, two main levels of combination are to be considered: the fusion of multiple features in Eq. (26) and the line/column-wise combination in Eq. (27). The first level of combination is interesting to investigate since its

objective is to fuse heterogeneous data described by belief functions in order
 535 to remove ambiguity in the association problem. In this pairwise combination
 of $m_{i,j}^p(\cdot)$ and $m_{i,j}^o(\cdot)$, two situations can be encountered: either the sources
 agree or highly conflict. When they disagree, the resulting conflict is important
 as it indicates an eventual mis-association due to a small variation of position
 but a large relative direction of motion and *vice versa*. Under the reliability
 540 hypothesis of each source, the conflict raised at this first fusion level informs
 about two objects that do not match. The conflict distribution is evaluated on
 the example depicted in Fig. 12 according to the combination rules presented
 in Section 3.4.

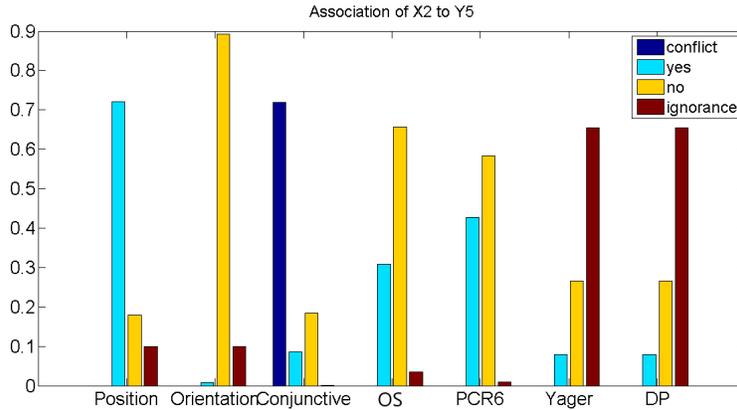


Figure 15: Belief expressed in the association of X_2 to Y_5 (from Fig. 12) by the standalone sources and their fusion.

In Fig. 15, the four masses expressing conflict, association (yes), non-association
 545 (no) and ignorance of the association of X_2 to Y_5 with respect to both sources
 and the results of five combination rules are displayed. It shows the high conflict
 due to the fact that X_2 and Y_5 are highly close whereas their motion
 is in complete opposition. The conflict appearing in the conjunctive combination
 is relatively high (0.719). The non-association mass expressed by the
 550 orientation source being high, the combined non-association mass (in normalized/conjunctive)
 is higher than the association mass. This also applies for the other combination rules.
 The only difference resides in the ignorance and con-

flict masses. Basically, both Yager and Dubois and Prade's (DP) give the same output because each source expresses its belief only on three focal elements: the association hypothesis of two particular objects, their non-association and the ignorance. Therefore, the disjunction $\{y\} \cup \{n\}$ is equivalent to the ignorance. Both Yager and DP induce less specific *bbas*. Thus, it is more interesting to use either Dempster's normalization or a more refined distribution such as PCR6.

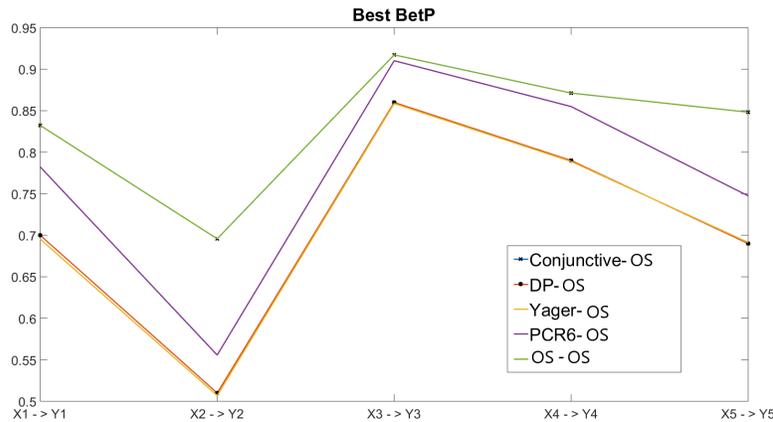


Figure 16: Combination rule evaluation at the pairwise fusion: pignistic probabilities corresponding to the best hypothesis.

In Fig. 16, the best pignistic probabilities corresponding to each correct association are displayed for the use case illustrated in Fig. 12. The reason for which these results are demonstrated is to evaluate the pairwise combination (first fusion level) on the position and orientation sources. For this first combination, all operators are used whereas for the second level, only the the normalized rule is used. It can be noticed that the resulting BetPs vary differently with respect to the chosen operator. For instance, Yager and DP's impact on the specificity is apparent as it has the lowest BetPs. However, Dempster's rule has quite assertive probabilities since they mostly vary between 0.7 and 0.9. PCR6 resulting probabilities remain approximately in the range of relevant values. Fig. 16 demonstrates that normalized-based pairwise combination has the largest confidence in the chosen hypotheses compared to the rest of the opera-

tors. Finally, a comparison is done between Dempster’s rule and PCR6 on the complete **Seq. 17** in Fig. 17. It can be noticed that the normalized combination maximizes the confidence in the best hypotheses. Such operation can facilitate the optimization step in decision-making by avoiding hazardous decisions and

575 can guarantee a robust association regardless the matching algorithm.

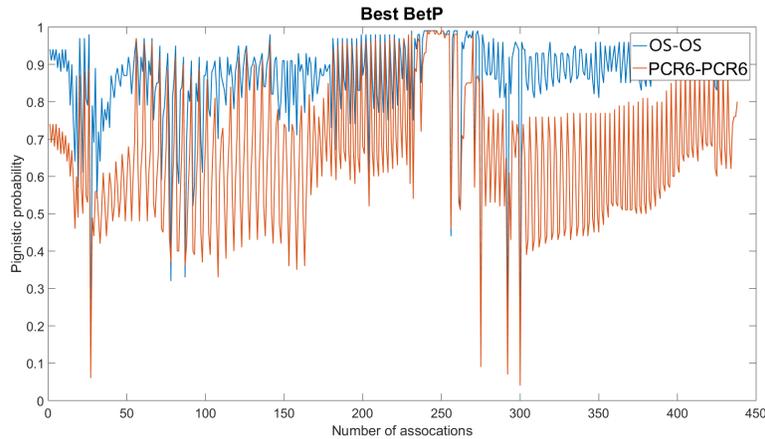


Figure 17: Best Pignistic Probabilities according to two combination strategies (OS: Orthogonal Sum, PCR6) on KITTI **Seq. 17**.

6. Conclusion and Perspectives

This paper tackles the problem of Multiple Object Association (MOA) in the Belief Function framework for Intelligent Vehicles (IV). It proposes three main contributions: the first one is based on an evidential multi-feature fusion

580 strategy to treat the MOA issue in cluttered environments. The approach is based on two heterogeneous sources defined by the position and size of dynamic objects as well as their direction of motion in the scene. This strategy provides a complementary information to the position source, commonly used, in order to ensure a distinction of close objects which could belong to the same category

585 (pedestrians, cars, etc.) but having different trajectories. Hence, this information removes ambiguity in decision-making. The evidence raised by each feature (position and orientation) is defined by independent mass models to generate

two separate sources. The second contribution relies on the selection and study of the appropriate mass models w.r.t. to the available sensor data. Hence, this paper first provides a recall of the most used mass models and discusses the specialized source model finally retained. The model identification process is also depicted. A credal fusion combines the position and orientation data. For this fusion, several rules are evaluated according to their rate of true associations on three complex sequences provided by the KITTI database. The proposed approach performed better than the individual sources as they provide different characteristics related to an object's compartment. The fusion, therefore, merges the evidence provided by each source to solve ambiguities. The third major contribution is the evaluation of the proposed strategy on an important and influential set of real data in the field of IV. Commonly, belief function-based MOA algorithms are validated on limited data sets. In this case, the paper stands out by the extensive objective validation it highlights. The results have also demonstrated the impact of the feature fusion in the association of dynamic objects to ensure a robust perception in the IV context.

Several perspectives could be considered. In the current work, a LiDAR provides the object direction of motion and a camera issues the position and size. An interesting research direction to investigate would be to define the object position using the LiDAR data and the direction of motion based on the image frames. Some preliminary tests of this configuration show very similar results compared to those of this paper. This investigation will be pursued for comparison purposes. Secondly, taking account of additional features related to the objects in the scene in order to be robust w.r.t. the large variety of driving scenarios is a necessary research direction. For instance, situations dealing with objects of different natures, kinematics, shapes, etc. require to consider valuable object-related information in the association process. Another future work will be to implement tracking to insure a complete supervision of complex environments. Finally, this algorithm will be tested on the experimental vehicle ARTEMIPS within the laboratory to validate its efficiency in a real-time multi-sensor data fusion framework.

References

- 620 [1] D. L. Hall, *Mathematical Techniques in Multisensor Data Fusion*, Artech House, Inc., Norwood, MA, USA, 1992.
- [2] D. Reid, An algorithm for tracking multiple targets, *IEEE Transactions on Automatic Control* 24 (6) (1979) 843–854. doi:10.1109/TAC.1979.1102177.
- 625 [3] S. Blackman, *Multiple-target Tracking with Radar Applications*, Radar Library, Artech House, 1986.
- [4] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [5] A. P. Dempster, A generalization of bayesian inference, *Journal of the Royal*
630 *Statistical Society* 30 (2) (1968) 205–247.
- [6] P. Smets, R. Kennes, The transferable belief model, *Artificial Intelligence* 66 (2) (1994) 191–234. doi:10.1016/0004-3702(94)90026-4.
- [7] D. Gruyer, S. Demmel, V. Magnier, R. Belaroussi, Multi-hypotheses tracking using the dempster-shafer theory. application to ambiguous road context., *Information Fusion* 29 (2016) 40–56. doi:10.1016/j.inffus.2015.10.001.
- 635 [8] M. Rombaut, V. Berge-Cherfaoui, Decision making in data fusion using Dempster-Shafer’s theory, in: *13th IFAC Symposium on Intelligent Components and Instrumentation for control Applications*, Annecy, France, 1997.
- 640 [9] M. Rombaut, Decision in multi-obstacle matching process using Dempster-Shafer’s theory, in: *International Conference on Advances in Vehicle Control and Safety*, Amiens, France, 1998, pp. 63–68.
- [10] C. Royère, D. Gruyer, V. Cherfaoui, Data association with believe theory, in: *International Conference on Information Fusion (FUSION)*, Vol. 1,

- 645 Paris, France, 2000, pp. TUD2/3–TUD2/9 vol.1. doi:10.1109/IFIC.
2000.862698.
- [11] B. Mourllion, D. Gruyer, C. Royère, S. Théroude, Multi-hypotheses tracking algorithm based on the belief theory, in: International Conference on Information Fusion (FUSION), Philadelphia, PA, USA, 2005, pp. 922–929.
650 doi:10.1109/ICIF.2005.1591957.
- [12] D. Mercier, E. Lefèvre, D. Jolly, Object association with belief functions, an application with vehicles, *Information Sciences* 181 (24) (2011) 5485–5500. doi:10.1016/j.ins.2011.07.045.
- [13] T. Denœux, N. El Zoghby, V. Cherfaoui, A. Jouglet, Optimal object association in the dempster-shafer framework, *IEEE Transactions on Cybernetics*
655 44 (22) (2014) 2521–2531. doi:10.1109/TCYB.2014.2309632.
- [14] J. Daniel, J. P. Lauffenburger, Multi-object association decision algorithms with belief functions, in: International Conference on Information Fusion (FUSION), Singapore, 2012, pp. 669–676.
- 660 [15] J. Daniel, J.-P. Lauffenburger, Fusing navigation and vision information with the transferable belief model: Application to an intelligent speed limit assistant, *Information Fusion* 18 (2014) 62–77. doi:10.1016/j.inffus.2013.05.013.
- [16] B. Ristic, P. Smets, Global cost of assignment in the tbm framework for association of uncertain id reports, *Aerospace Science and Technology* 11 (4)
665 (2007) 303–309, COGIS '06. doi:10.1016/j.ast.2006.10.008.
- [17] W. Rekik, S. L. Hégarat-Masclé, R. Reynaud, A. Kallel, A. B. Hamida, Dynamic object construction using belief function theory, *Information Sciences* 345 (2016) 129 – 142. doi:10.1016/j.ins.2016.01.047.
- 670 [18] R. O. Chavez-Garcia, O. Aycard, Multiple sensor fusion and classification for moving object detection and tracking, *IEEE Transactions on Intelligent*

Transportation Systems 17 (2) (2016) 525–534. doi:10.1109/TITS.2015.2479925.

- [19] S. Hachour, F. Delmotte, D. Mercier, A robust credal assignment solution
675 based on the generalized bayes' theorem, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 25 (06) (2017) 947–971. doi:10.1142/s0218488517500416.
- [20] Y. Dan, J. Hongbing, G. Yongchan, A robust D-S fusion algorithm for
680 multi-target multi-sensor with higher reliability, Information Fusion 47 (2019) 32–44. doi:10.1016/j.inffus.2018.06.009.
- [21] M. Boumediene, J. P. Lauffenburger, J. Daniel, C. Cudel, A. Ouamri, Multi-roi association and tracking with belief functions: Application to traffic sign recognition, IEEE Transactions on Intelligent Transportation Systems 15 (6) (2014) 2470–2479. doi:10.1109/TITS.2014.2320536.
- [22] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving?
685 the kitti vision benchmark suite, in: Conference on Computer Vision and Pattern Recognition (CVPR), Rhode Island, USA, 2012. doi:10.1109/CVPR.2012.6248074.
- [23] H. Laghmar, T. Laurain, C. Cudel, J. P. Lauffenburger, 2.5d evidential
690 grids for dynamic object detection, in: International Conference on Information Fusion (FUSION), Ottawa, Canada, 2019, pp. 1–7.
- [24] H. Laghmar, C. Cudel, J. P. Lauffenburger, M. Boumediene, On the information selection for optimal data association, in: International Conference on Information Fusion (FUSION), Xi'an, China, 2017, pp. 1–8.
695 doi:10.23919/ICIF.2017.8009706.
- [25] T. Dencœux, A k-nearest neighbor classification rule based on dempster-shafer theory, IEEE Transactions on Systems, Man, and Cybernetics 25 (5) (1995) 804–813. doi:10.1109/21.376493.

- [26] E. Lefevre, O. Colot, P. Vannoorenberghe, Belief function combination and conflict management, *Information Fusion* 3 (2) (2002) 149–162. doi:10.1016/S1566-2535(02)00053-2.
- [27] A. Martin, C. Osswald, Toward a combination rule to deal with partial conflict and specificity in belief functions theory, in: *International Conference on Information Fusion (FUSION)*, Québec, Canada, 2007, p. 1190. doi:10.1109/ICIF.2007.4408007.
- [28] J.-B. Yang, D.-L. Xu, Evidential reasoning rule for evidence combination, *Artificial Intelligence* 205 (2013) 1 – 29. doi:10.1016/j.artint.2013.09.003.
- [29] R. Yager, On the dempster-shafer framework and new combination rules, *Information Sciences* 41 (1987) 93–137. doi:10.1016/0020-0255(87)90007-7.
- [30] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, *Computational Intelligence* 4 (3) (1988) 244–264. doi:10.1111/j.1467-8640.1988.tb00279.x.
- [31] F. Smarandache, J. Dezert, Information fusion based on new proportional conflict redistribution rules, in: *International Conference on Information Fusion (FUSION)*, Vol. 2, 2005, p. 8. doi:10.1109/ICIF.2005.1591955.
- [32] P. Smets, Constructing the pignistic probability function in a context of uncertainty, in: *Uncertainty in Artificial Intelligence*, Vol. 10, 1990, pp. 29–39. doi:10.1016/B978-0-444-88738-2.50010-5.
- [33] H. Laghmara, M. Boumediene, C. Cudel, J. P. Lauffenburger, Evidential object association using heterogeneous sensor data, in: *International Conference on Information Fusion (FUSION)*, Cambridge, UK, 2018, pp. 1–8. doi:10.23919/ICIF.2018.8455711.