Fusion at Detection Level for Frontal Object Perception

R. Omar Chavez-Garcia, Trung Dung Vu, Olivier Aycard

Fusion at Detection Level for Frontal Object Perception

R. Omar Chavez-Garcia[1], Trung-Dung Vu[2] and Olivier Aycard[3]

Abstract— **Intelligent vehicle perception involves the correct detection and tracking of moving objects. Taking into account all the possible information at early levels of the perception task can improve the final model of the environment. In this paper, we present an evidential fusion framework to represent and combine evidence from multiple lists of sensor detections. Our fusion framework considers the position, shape and appearance information to represent, associate and combine sensor detections. Although our approach takes place at detection level, we propose a general architecture to include it as a part of a whole perception solution. Several experiments were conducted using real data from a vehicle demonstrator equipped with three main sensors: lidar, radar and camera. The obtained results show improvements regarding the reduction of false detections and mis-classifications of moving objects.**

I. INTRODUCTION

Intelligent Vehicle Perception (IVP) relies on sensor data to model the static and moving parts of the environment. IVP is composed of two main tasks: simultaneous localization and mapping (SLAM) deals with modeling static parts; and detection and tracking moving objects (DATMO) is responsible for modeling dynamic parts of the environment. In SLAM, when vehicle location and map are unknown the vehicle generates a map of the environment while simultaneously localizing itself in the map given all the measurements from its sensors. DATMO aims at detecting and tracking the moving objects surrounding the vehicle in order to predict their future behaviors [1], [2].

Usually, the tracking process assumes that its inputs correspond uniquely to moving objects. However, in most of the real outdoor scenarios, inputs include non-moving detections, such as noisy measurements or static obstacles. Sensors technical limitations contribute to these impressions. Accurate detection of moving objects is a critical aspect of a moving object tracking system. Therefore, many sensors are part of a common intelligent vehicle system.

Multiple sensor fusion has been a topic of research since long; the reason is the need to combine information from different views of the environment to obtain a more accurate model. This is achieved by combining redundant and complementary measurements of the environment. Inside the DATMO component, fusion can be performed at two levels [2]. At object detection level, sensor processes provide lists of moving object detections, then these lists are combined to get an enhanced list. At tracking level, lists of tracks of moving objects are fused to produce an enhanced list of tracks.

Classification of moving objects is needed to determine the possible behavior of the objects surrounding the vehicle, and it is usually performed at tracking level. Knowledge about the class of moving objects at detection level can help to improve their tracking, reason about their behavior and decide what to do according to their nature [2], [3].

Labayrade et al. presented a fusion technique between laser and stereo vision for obstacles detection [4]. This technique is based on stereo vision segmentation and lidar data clustering.Redundant positions in both sensor detections are considered real moving objects. Detections having no matching counterparts are taken as false alarms and are ignored. We believe that this is a strong assumption and that position information could not be enough to decide if an object is real.

Fayad et al. have proposed an evidential fusion technique based on the Dempster-Shafer (DS) theory [5]. This technique is a mixture of object detection and tracking level fusion. Their work focuses only on the detection of pedestrians using multiple sensors by maintaining an score for each detection. Although the results are promising, this work only considers class information to perform object fusion, leaving out location information. Moreover, the extension to detect multiple moving objects classes is not straightforward.

Fusion at object detection level can enrich the object representation, allowing the tracking process to rely on this information to make better association decisions and obtain better object estimates. However, when combining different sensor inputs, we must take into account the classification precision of each sensor [1], [2].

In this paper, we propose a fusion approach at detection level based on DS theory. We use all the detection information provided by the sensors (i.e., position, shape and class information) to build a composite object representation. Given several lists of object detections, the proposed approach performs an evidential data association method to decide which detections are related and then fuses their representations. We use lidar and radar sensor to provide an approximate detection's position; and we use shape, relative speed and visual appearance features to provide a preliminary evidence distribution of the class of the detected objects. The proposed method includes uncertainty from the sensor detections without discarding non-associated objects. Multiple objects of interest are detected: *pedestrian, bike,*

*This work was also supported by the European Commission under interactIVe (http://www.interactive-ip.eu), a large scale integrating project part of the FP7-ICT for Safety and Energy Efficiency in Mobility. The authors would like to thank all partners within interactIVe for their cooperation and valuable contribution.

Laboratoire d'Informatique de Grenoble, University of Grenoble1, Grenoble, France {[1]ricardo.chavez-garcia, [2]trung-dung.vu, [3]olivier.aycard}@imag.fr

car and *truck*. Our method takes place at an early stage of DATMO component but we present it inside a complete real-time perception solution.

In order to evaluate our approach we used real data from highways and urban areas. The data were obtained using a vehicle demonstrator from the interactIVe(Accident Avoidance by Active Intervention for Intelligent Vehicles) European project. Our experiments aim at evaluating the degree of improvement in DATMO results when early combination of class information is performed.

The rest of the paper is organized as follows. Next section describes the vehicle demonstrator and its sensor configuration. Section III reviews some concepts of the DS theory. In Section IV, we define our fusion framework at detection level. The implementation of this fusion framework is done using the architecture define in Section V. Experimental results are shown in Section VI. Finally, Section VII presents the conclusions.

II. Vehicle demonstrator

The CRF vehicle demonstrator we used is part of the interactIVe European project. The demonstrator is a Lancia Delta car equipped from factory with electronic steering systems, two ultrasonic sensors located on the side of the front bumper, and with a front camera located between the glass and the central rear mirror. Moreover, the demonstrator vehicle has been equipped with an scanning laser (lidar) and a mid-range radar on the front bumper for the detection of obstacles ahead, as depicted in Figure 1.

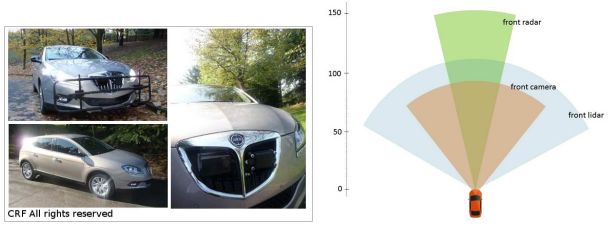


Fig. 1. Left: images of the CRF vehicle demonstrator. Right: Field of view of the three frontal sensors.

III. Dempster-Shafer theory background

DS theory is considered a generalization of the Bayesian theory of subjective probability. Whereas the Bayesian theory requires probabilities for each question of interest, DS theory allows us to base degrees of belief for one question on probabilities for a related question [6]. This theory is highly expressive, allows to represent different levels of ignorance, does not require prior probabilities, and manage conflict situations when opposite evidence appears.

DS theory represents the world in a set of mutually exclusive propositions known as the frame of discernment (Ω). It uses belief functions to distribute the evidence about the propositions over 2^{Ω}. The distribution of mass beliefs is done by the function $m : 2^{\Omega} \rightarrow [0, 1]$, also known as the Basic Belief Assignment (BBA), which is described in (1). DS representation allows scenarios where there is uncertain evidence about all the proposition. Besides, a BBA can support any proposition $A \subseteq \Omega$ without supporting any sub-proposition of A, which allows to express partial knowledge.

$$m(\emptyset) = 0; \sum_{A \subseteq \Omega} m(A) = 1 \tag{1}$$

We can represent the evidence from two sources as belief functions over the same frame of discernment. A combination rule takes these two belief distributions and combine them into a new one. Dempster's rule of combination is one of the most widely used [6]. It assumes independence and reliability of both sources of evidence:

$$\begin{aligned} m_{12}(A) =& \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K_{12}}; \; A \neq \emptyset \\ K_{12} =& \sum_{B \cap C = \emptyset} m_1(B) m_2(C) \end{aligned} \tag{2}$$

where K_{12} is known as the degree of conflict. Dempster's rule analyses each piece of evidence to find conflict and uses it to normalize the masses in the set. However, in scenarios with high conflict values this normalization leads to counter intuitive scenarios. A possible solution is to avoid this normalization by moving the conflict evidence K_{12} to all the possible elements of the frame of discernment Ω.

$$\begin{aligned} m_{12}(A) =& \sum_{B \cap C = A} m_b(B) m_c(C); \; A \neq \emptyset \\ K_{12} =& \sum_{B \cap C = \emptyset} m_b(B) m_c(C) \\ m_{12}(\Omega) =& m'_{12}(\Omega) + K_{12} \end{aligned} \tag{3}$$

where $m'_r(\Omega)$ is the combined evidence for the ignorance hypothesis. This modified Dempster's rule is known as Yager's combination rule.

IV. Fusion at detection level

Our work proposes a sensor fusion framework placed at detection level. Although this approach is presented to work with three main sensors, it can be extended to work with more sources of evidence. Figure 2 shows the general architecture of the proposed fusion approach. The inputs of this method are several lists of detected objects. Each detection is represented by its position and an evidence distribution of its class represented as a BBA. The reliability of the sources of evidence is encoded inside the BBAs. Class information is obtained from the shape, relative speed and visual appearance of the detections. The final output of the fusion method comprises a fused list of object detections, represented by a composite representation that includes: position, shape and an evidence distribution of class hypotheses.

A. Object detection representation

Usually, object detections are represented by their position and shape features. We believe that class information can be important to consider at detection level. However, at this level there is not enough certainty about the class of the object. Hence, keeping only one class hypothesis per

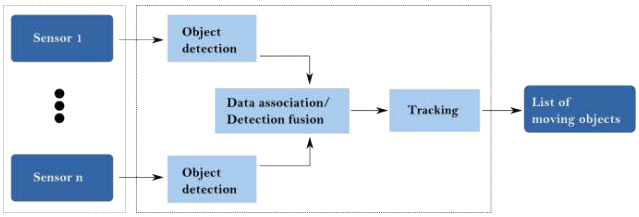


Fig. 2. Schematic of the proposed fusion architecture.

detection disables the possibility of rectifying a premature decision.

Our composite object representation is defined by two parts. First, it includes position and shape information in a two dimensional space. Second, it includes an evidence distribution m_c for all possible class hypotheses 2^{Ω_c}, where Ω_c is the frame of discernment representing the classes of moving objects of interest.

B. Data association

Let us consider two sources of evidence S_1 and S_2. Each of these sources provides a list of detections denoted by $A = a_1, a_2, .., a_b$ and $B = b_1, b_2, ..b_n$ respectively. In order to combine the information of these sources we need to find the associations between the object detections in A and B. All possible associations can be expressed as a matrix $M_{A,B}$ of magnitude $|AxB|$, where each cell represents the evidence m_{a_i,b_j} about the association of the elements a_i and b_j for $i < |A|$ and $j < |B|$. We can define three propositions regarding the association of the detections in A and B:

- $P(a_i, b_j) = 1$: if a_i and b_j are the same object
- $P(a_i, b_j) = 0$: if a_i and b_j are not the same object
- $P(a_i, b_j) = \Omega$: ignorance about the association of the detections a_i and b_j

Let us define $\Omega = \{1, 0\}$ as the frame of discernment of each m_{a_i,b_j}, where $\{1\}$ means that detection a_i and b_j belong to the same object, and $\{0\}$ otherwise. Therefore, $m_{a_i,b_j}(\{1\})$ and $m_{a_i,b_j}(\{0\})$ quantify the evidence supporting the proposition $P(a_i, b_j) = 1$ and $P(a_i, b_j) = 0$ respectively; and $m_{a_i,b_j}(\{1, 0\})$ stands for the ignorance, i.e., the evidence that can not support the other propositions. The three different propositions can be addressed by representing the similarity of the detections in A and B. This means, m_{a_i,b_j} can be defined based on similarity measures between detections a_i and b_j.

Sensors S_1 and S_2 can provide detections of different kind. These detections can be represented by a position, shape or appearance information, such as class. Hence, m_{a_i,b_j} has to be able to encode all the available similarity information. Let us define m_{a_i,b_j} in terms of its similarity value as follows.

$$\begin{aligned} m_{a_i,b_j}(\{1\}) &= \alpha_{i,j} \\ m_{a_i,b_j}(\{0\}) &= \beta_{i,j} \\ m_{a_i,b_j}(\{1,0\}) &= 1 - \alpha_{i,j} - \beta_{i,j} \end{aligned} \tag{4}$$

where $\alpha_{i,j}$ and $\beta_{i,j}$ quantify the evidence supporting the singletons in Ω for the detections a_i and b_j, i.e., the similarity measures between them.

We can define m_{a_i,b_j} as the fusion of all possible similarity measures to associate detections a_i and b_j. Therefore, we can assume that individual masses of evidence carry specific information about these two detections. Let us define m^p as the evidence measure about the position similarity between detections in A and B provided by sources S_1 and S_2 respectively; and m^c as the evidence measure about the appearance similarity. Following the analysis made in Section III, we used Yagers's combination rule to represent m_{a_i,b_j} in terms of $m^p_{a_i,b_j}$ and $m^c_{a_i,b_j}$ as follows:

$$\begin{aligned} m_{a_i,b_j}(A) &= \sum_{B \cap C = A} m^p_{a_i,b_j}(B) m^c_{a_i,b_j}(C) \\ K_{a_i,b_j} &= \sum_{B \cap C = \emptyset} m^p_{a_i,b_j}(B) m^c_{a_i,b_j}(C) \\ m_{a_i,b_j}(\{\Omega\}) &= m'_{a_i,b_j}(\{\Omega\}) + K_{a_i,b_j} \end{aligned} \tag{5}$$

where $m^p_{a_i,b_j}$ and $m^c_{a_i,b_j}$ represent the evidence about the similarity between detections a_i and b_j taking into account the position information and the class information, respectively.

Once the matrix $M_{A,B}$ is built, we can analyze the evidence distribution m_{a_i,b_j} for each cell to decide if there is an association ($m_{a_i,b_j}(\{1\})$), there is not ($m_{a_i,b_j}(\{0\})$) or we have not enough evidence to decide ($m_{a_i,b_j}(\{1, 0\})$), which is probably due to noisy detections. In the next sections we will describe how to calculate the fused evidence distributions using similarity evidence from the detections.

The fused representation is obtained by combining the evidence distributions between the associated objects by applying the combination rule from (3). This representation is passed as an input to the tracking stage to be considered in the motion model estimation of the moving objects. Non-associated objects detections are passed as well expecting to be deleted by the tracking process if they are false detections or to be verified as real objects in case that more evidence confirms these detections.

C. Position similarity

According to the position of two detections a_i and b_j, we encode their similarity evidence in $m^p_{a_i,b_j}$. Based on their positions we can define the function d_{a_i,b_j} as a distance function that satisfies the properties of a pseudo-distance metric. We choose Mahalanobis distance due to its ability to include the correlations of the set of distances [7]. Therefore, a small value of d_{a_i,b_j} indicates that detections a_i and b_j are part of the same object; and a large value indicates the opposite. Hence, the BBA for $m^p_{a_i,b_j}$ is described as follows:

$$\begin{aligned} m^p_{a_i,b_j}(\{1\}) &= \alpha f(d_{a_i,b_j}), \\ m^p_{a_i,b_j}(\{0\}) &= \alpha(1 - f(d_{a_i,b_j})), \\ m^p_{a_i,b_j}(\{1,0\}) &= 1 - \alpha, \end{aligned} \tag{6}$$

where $\alpha \in [0, 1]$ is an evidence discounting factor and $f(d_{a_i,b_j}) \rightarrow [0, 1]$. The smaller the distance, the larger value given by function f.

D. Class dissimilarity

Contrary to the evidence provided by position, class information does not give evidence that supports the proposition $P(a_i, b_j) = 1$. This means that even if two detections are identified with the same class, one can not affirm that are the same object. This is due to the fact that there can be multiple different objects of the same class, e.g., in a real driving scenario many cars or pedestrians can appear. However, it is clear that if two detections have different class it is more likely that they belong to different objects. Hence, we use the class information to provide evidence about the dissimilarity of detections and place it in $m^c_{a_i,b_j}$. The frame of discernment for the class evidence distribution is the set Ω_c of all possible classes. The frame of discernment for detections association is Ω and was described in Section IV-B. Hence, we propose to transfer the evidence from in Ω_c to Ω as follows.

$$\begin{aligned} m^c_{a_i,b_j}(\{1\}) =& 0 \\ m^c_{a_i,b_j}(\{0\}) =& \sum_{A\cap B=\emptyset} m^c_{a_i}(A) m^c_{b_j}(B), \\ & \forall A, B \subset \Omega_c \\ m^c_{a_i,b_j}(\{1,0\}) =& 1 - m^c_{a_i,b_j}(0) \end{aligned} \tag{7}$$

which means that we fuse the mass evidences where no common class hypothesis is share between detections in lists A and B. $m^c_{a_i}$ and $m^c_{b_j}$ represent the BBAs for class hypotheses of detections in lists A and B. However, as we have no information about the possible relation of detections with the same class, we place the rest of the evidence in the ignorance hypothesis $\{1, 0\}$.

V. Frontal Object Perception Application

Figure 3 shows the general architecture of a frontal object perception application developed in the interactIVe project for the vehicle demonstrator described in Section II. The purpose of this architecture is to detect, classify and track a set of possible objects (pedestrian, bike, car and truck) in front of the vehicle demonstrator. Our proposed work is embedded in the fusion component of this architecture. It takes object detections from three sensors: radar, lidar and camera. Although there are three sensors, only two lists of object detections are provided as inputs to the fusion module.

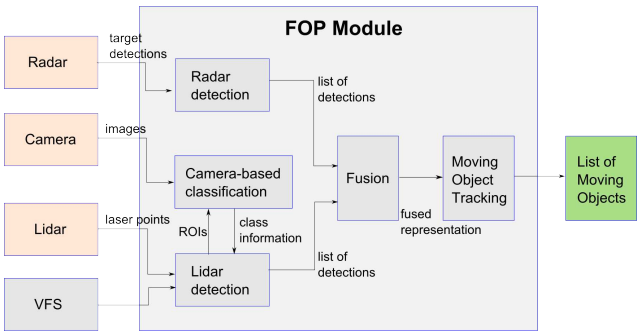


Fig. 3. General architecture of the FOP module for the CRF demonstrator.

A. Lidar object detection

Raw lidar scans and vehicle state information are processed to built a static map and detect moving objects. We employed the probabilistic grid-based approach presented in [1] to incrementally integrate discrete lidar scans into a local 2D occupancy grid. Inconsistencies through time in the occupancy grid allow the method to detect moving obstacles. The points observed in free space are classified as moving whereas the rest are classified as static. Using a clustering process we identify groups of points that could describe moving objects.

Once the clusters of possible moving objects are built, a bounding-box representation is drawn for each cluster, allowing to extract a visible shape-based description of each moving object. Shape information allows to have a first clue of the class of the detected object. First of all, we need to define the frame of discernment for the class hypotheses $\Omega_c = \{pedestrian, bike, car, truck\}$. According to the size of the object, we assign evidence to certain class hypotheses. We follow a fix-size model to do so. However, no precise classification decision can be made due to the temporary visibility of the moving objects. If the width of the bounding box is less than a threshold ω_w we can think the object is a pedestrian or a bike but we can not be sure of the real size of the object. If the width of the object is greater than ω_w the object is less likely to be a pedestrian or bike, but it can be either a car or a truck. We follow the initial evidence mass distribution and the discounting factors proposed in [8] to incorporate uncertainty about the classification evidence.

B. Camera based object classification

Lidar processing detection provides a rough classification of the detected moving objects. This classification relies on the visible shape of the detection which is a strong assumption in a highly dynamic environment. We believe that an appearance based classification could provide more certainty about the class of the detected objects. Therefore, we use the lidar detections to generate regions of interest (ROIs) in the camera images. Moreover, lidar detections give a better estimation of the real shape of the object. The ROIs are taken by vehicle and pedestrian classifiers to perform the camera-based classification. A modified version of histogram of oriented gradients (called sparse-HOG) features, which focus on important areas of the samples, powers the pedestrian and vehicle visual descriptor at training and detection time. Given computed descriptors for positive and negative samples, we use the discrete Adaboost approach proposed in [9] to train the vehicle and pedestrian classifiers. Its trade-off between performance and classification precision makes it suitable for real-time requirements.

Pedestrian and Vehicle classifiers are used to built another mass evidence distribution with the same frame of discernment described in Section V-A. Following the same basic belief assignment proposed in [8] we built a camera-based evidence distribution over the frame of discernment $\Omega_c = \{pedestrian, bike, car, truck\}$. Afterwards, using (3) we combine camera based class distribution with the lidar

based distribution mentioned in Section V-A to obtain our first input for our fusion approach.

C. Radar target detection

Radar sensors have a good range resolution and a crude azimuth estimation. They usually provide an estimation of the position of moving targets and their relative speed. We use this estimation to built a second list of object detections as we did for the lidar data. However, the lack of shape information makes difficult to have clues about the object class at detection level. Although an estimate, the relative speed can give clues about the nature of the object. An detection with a high relative speed is most likely to be a motor-based vehicle such as car or truck. Nonetheless, no class assumption can be made about a low-speed object. Following the same idea and frame of discernment from Section V-A, we built a BBA by placing evidence in the vehicle hypothesis ($\{car, truck\}$) when the relative speed of a detection is greater than a speed threshold s_{rel} (fixed a priori). In other cases, we put the evidence in the ignorance hypothesis $\{\Omega\}$. Discounting factors are applied to take into account the uncertainty from radar detections.

D. Fusion considerations

Once we have performed moving object detection using lidar processing, the proposed approach obtains a preliminary description of the object position and object class encoded in m^c_{lidar}. Afterwards, taking advantage of the accuracy of the ROIs obtained by lidar and executing the camera based classifiers, a second evidence class distribution m^c_{camera} is obtained. These two evidence distributions are combined using (3) to form m^c_a.

Radar processing provides already a list of detections identified by their position and relative speed. Following the method describe in Section V-C, we built the class distribution for radar detections m^c_b. Finally, both lists of object representations are processed in order to identify their associations and fuse their evidence distributions using (5), (6) and (7).

E. Moving object tracking

Moving object tracking has to be performed in order to deliver the final output of a perception system. We follow the moving object tracking approach proposed by Vu [1]. We adapted this work to represent not only the lidar measurements but the composite representation obtained by our proposed fusion approach. Tracking mechanism interprets the composite representations sequence by all the possible hypotheses of moving object trajectories over a sliding window of time. Generated object hypotheses are then put into a top-down process taking into account all object dynamics models, sensor models and visibility constraints. We use the class evidence distribution to reduce the number of generated hypotheses by considering only class hypotheses with the highest mass evidence in 2^{Ω_c}.

TABLE I
VEHICLE (CAR AND TRUCK) MIS-CLASSIFICATIONS OBTAINED BY THE FUSION APPROACHES.

Dataset	Number of vehicles	Number of vehicle mis-classifications	
		Tracking level	Detection level
highway 1	35	7	4
highway 2	42	6	5
urban 1	82	19	10
urban 2	120	23	8

VI. EXPERIMENTS AND RESULTS

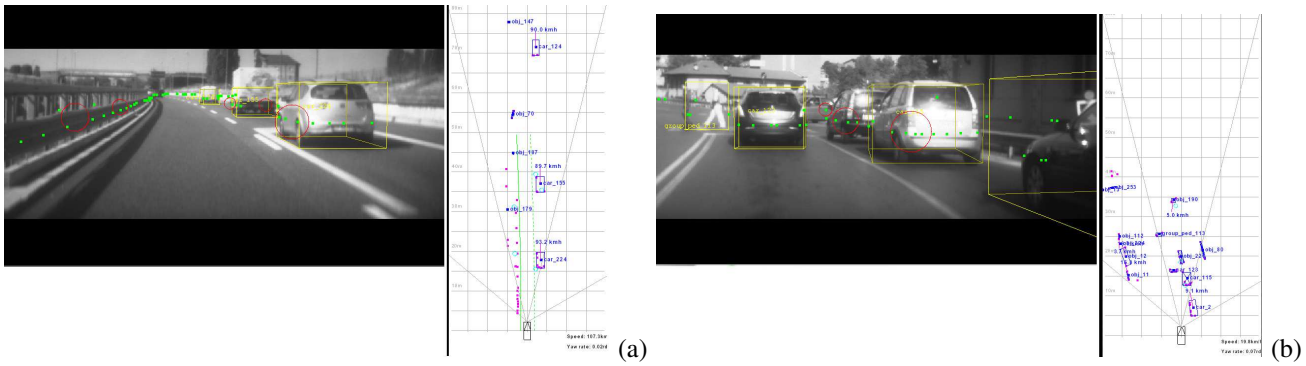
Using the vehicle demonstrator described in Section II, we gathered four datasets from real scenarios: two datasets from urban areas; and two datasets from highways. As a comparison approach we use our fusion approach at tracking level described in [8] which takes as inputs the same datasets. The goal of these experiments was to analyze the degree of improvement achieved by early inclusion of class information within the DATMO component.

We follow the general architecture from Section V to build our a complete perception system and test it using the gathered datasets. Among the 2D position state for each object detection, we define the frame of discernment $\Omega_c = \{pedestrian, bike, car, truck\}$ for its evidence class distribution. Therefore, 2^Ω represents all the possible class hypothesis for each detection.

Figure 4 (a) shows three vehicles in front of the vehicle demonstrator. However, only two radar detections (from several spurious detections) are correct. In this situation, lidar based detection and camera based classification evidence placed in m^c_{lidar} and m^c_{camera} correctly complement the information about the farthest vehicle. Besides, moving object class is determined sooner than in the fusion approach at tracking level due to the early fused evidence about the object's class. False moving object detections are not deleted when fusion is performed, but they are passed to the tracking approach which will discard them after few non-associations.

Figure 4 (b) shows a cross road situation in a urban scenario. All the moving objects are detected but one car in the very front of the waiting line. Although the car is sensed by radar, there is not enough evidence from lidar detection and camera-based classification to verify its moving state. Moreover, the car is barely seen by lidar and few frames have passed to determine if it is moving or not. This car is consider a static unclassified object and appears in the top view. The car just behind this unrecognized car is as well consider static but it is identified and classified due to the previous detections that allowed to determine its moving nature.

Tables I and II show a comparison between the results obtained by the proposed fusion approach at detection level and our previous fusion approach at tracking level taking into account the mis-classifications of moving objects. Regarding the pedestrian classification, the obtained reduction in the number of mis-classifications does not seem to be as relevant



(a) (b)

Fig. 4. Frontal object perception results for (a) highway scenario and (b) urban area. Left side of each figure shows the image from camera sensor and the identified moving objects. Yellow boxes represent moving objects, red dots represent lidar hits and red circles represent radar detections. Right side of each figure shows the top view of the scene show in the image. Objects classes are shown by tags close to each object.

TABLE II

PEDESTRIAN AND BIKE MIS-CLASSIFICATIONS OBTAINED BY THE FUSION APPROACHES. HIGHWAY DATASETS CONTAIN ONLY VEHICLES.

Dataset	Number of pedestrians and bikes	Number of pedestrian mis-classifications	
		Tracking level	Detection level
urban 1	21	6	5
urban 2	23	8	6

TABLE III

MOVING OBJECTS FALSE DETECTIONS OBTAINED BY THE FUSION APPROACHES.

Dataset	Number of object false detections	
	Tracking level	Detection level
highway 1	7	4
highway 2	8	5
urban 1	18	10
urban 2	19	9

as the other results. However, the classification of moving objects (not only pedestrians) in our proposed approach takes in average less sensor scans than the fusion approach described in [8] due to the early integration of the knowledge about the class of the objects placed in m_a^c and m_b^c.

Table III shows the number of false detections obtained by the fusion at detection level and by the fusion approach at tracking level. In our experiments, a false detection occurs when a detection is identified as moving when it is not. This false detections occur due to noisy measurements and wrong object associations which are directly related to the lack detection data, e.g., position, size and class. The obtained results show that combining all the available information from detections at detection level reduces the number of mis-detections and therefore provides a more accurate list of objects to the tracking process, which ultimately improve the final result of the frontal object perception system. Furthermore, the implementation of our proposed fusion scheme complies with the real-time constrain required for real automotive applications.

VII. CONCLUSIONS

In this paper, we presented a multiple sensor fusion framework at detection level based on DS theory to represent class hypotheses, associate object detections and combine evidence from their position and appearance. Even if we use a specific set of sensors to feed our proposed fusion approach, it can be extended to include several sources of evidence. The proposed method includes uncertainty from the evidence sources and from the object classification.

Several experiments were conducted using datasets from real driving scenarios. We showed a quantitative comparison between the presented fusion approach at detection level and a fusion approach at tracking level. These experiments showed improvements in the reduction of mis-classifications and false detections of moving objects.

REFERENCES

[1] T.-D. Vu, "Vehicle Perception : Localization , Mapping with Detection , Classification and Tracking of Moving Objects," Ph.D. Thesis, University of Grenoble 1, 2009.

[2] Q. Baig, "Multisensor Data Fusion for Detection and Tracking of Moving Objects From a Dynamic Autonomous Vehicle," Ph.D. dissertation, University of Grenoble1, 2012.

[3] C. Wang, C. Thorpe, S. Thrun, M. Hebert, and H. Durrant-Whyte, "Simultaneous localization, mapping and moving object tracking," *The International Journal of Robotics Research*, vol. 26, no. 9, pp. 889–916, 2007.

[4] R. Labayrade, C. Royere, D. Gruyer, and D. Aubert, "Cooperative fusion for multi-obstacles detection with use of stereovision and laser scanner," *Autonomous Robots*, vol. 19, no. 2, pp. 117–140, 2005.

[5] F. Fayad and V. Cherfaoui, "Detection and Recognition confidences update in a multi-sensor pedestrian tracking system," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 2008, pp. 409–416.

[6] P. Smets, "The Transferable Belief Model for Belief Representation," vol. 6156, no. Drums Ii, pp. 1–24, 1999.

[7] A. Bellet, A. Habrard, and M. Sebban, "A Survey on Metric Learning for Feature Vectors and Structured Data," *CoRR*, vol. abs/1306.6, 2013.

[8] R. Chavez-Garcia, T.-D. Vu, O. Aycard, and F. Tango, "Fusion framework for moving-object classification," in *Information Fusion (FUSION), 2013 16th International Conference on*, 2013, pp. 1159–1166.

[9] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine learning*, vol. 37, no. 3, pp. 297–336, 1999.