Map-Aided Evidential Grids for Driving Scene Understanding

Marek Kurdej, Julien Moras, Philippe Bonnifait, Véronique Cherfaoui

Map-aided Evidential Grids for Driving Scene Understanding†

Marek Kurdej*, Julien Moras®, Véronique Cherfaoui*, Philippe Bonnifait*

***Abstract*— Evidential grids have recently shown interesting properties for mobile object perception since the Dempster–Shafer framework allow them to handle efficiently partial information which is a frequent situation when driving in complex urban areas. This article deals with a lidar perception scheme that is enhanced by geo-referenced maps used as an additional source of information in a multi-grid fusion framework. The paper presents the key stages of such a data fusion process. An adaptation of the conjunctive combination rule is presented to refine the analysis of the conflicting information. The method relies on temporal accumulation to make the distinction between stationary and moving objects, and applies contextual discounting for modelling information obsolescence. As a result, the method is able to better characterise the state of the occupied cells by differentiating moving objects, parked cars, urban infrastructure and buildings. Another output of this approach is the capability to separate the navigable space from the non-navigable one. Experiments carried out on real traffic conditions with an equipped car illustrate the performance of such an approach.**

***Index Terms*— dynamic fusion, geo-referenced maps, mobile perception, prior knowledge, evidential occupancy grid, autonomous vehicle**

I. Introduction

Autonomous driving has been an important challenge in recent years. Navigation and precise localisation aside, perception of nearby environment is an important characteristic of a self-drivable vehicle. The level of difficulty in autonomous driving increases in urban environments, where a good scene understanding makes the perception crucial. On the one hand, there are several reasons that make cities a demanding environment. Poor satellite visibility deteriorates the precision of GPS positioning. Objects' trajectories are hard to predict due to high variation in speed and direction. Also, the sheer number of mobile objects poses computational problems, e.g. for tracking algorithms.

On the other hand, more and more detailed and precise geographic databases become available. This source of information has not been well examined yet, hence our approach of incorporating prior knowledge from digital maps in order to improve perception scheme. A substantial amount of research has focused on the mapping problem for autonomous vehicles, e.g. Simultaneous Localisation and Mapping (SLAM) approach [2], but the use of maps for perception is still understudied.

In this article, we propose a new perception scheme for intelligent vehicle navigation. The information fusion method is based on Dempster–Shafer theory of evidence [3]. The principal innovation of the method is the use of meta-knowledge obtained from a digital map. The map is considered as an additional source of information on a par with other sources, e.g. sensors. We study the advantage of including prior knowledge into an embedded perception system of an autonomous car. To model the vehicle environment, our approach uses multiple 2D occupancy grids [4], or precisely their evidential version adaptation [5].

Digital maps have been used as prior knowledge for already a few years [6]. The use of navigable road maps for localisation [7], [8], perception [9], navigation and driving [10] was proved to be of high value. Not only highly accurate and precised specialised maps, but also those which are open-sourced have been successfully employed, e.g., OpenStreetMap [11], [12]. A considerable amount of research works have contributed to the map updating problem [13].

Our method aims to capture complex and dynamic vehicle environments, so that they can be used as world representations for safe navigation. In particular, it is important to make the vehicle able to distinguish static objects attached to infrastructure (e.g. road sign) from mobile road users (e.g. cars, cyclist, pedestrian). In this last category (called "movable" in the following), the perception mechanism can be enhanced in order to make the difference between stopped and moving objects. Indeed, if a movable object has been detected on the side of the carriageway, it can be a pedestrian and the vehicle may reduce its speed in anticipation of a possible collision. Another objective of the proposed scheme is to characterize the free space nearby the vehicle by using exteroceptive sensors and refine this information by highlighting, thanks to a map of the area, the navigable space on which the vehicle can plan a safe path.

This paper describes a unified approach to a variety of problems in spatial representation using the Dempster–Shafer theory of evidence. The theory of evidence was not combined with occupancy grids until recently to build environment maps for robot perception [5]. Only recent works take advantage of the theory of evidence in the context of mobile perception [14]. There is also some research on efficient probabilistic and 3-dimensional occupancy grids [15]. Some authors have also used a laser range scanner as an exteroceptive source of information [14]. Some works use 3D city model as a source of prior knowledge for localisation and vision-based perception [16], whereas our method uses maps for scene understanding. Geodata are also successfully used for mobile navigation [17].

† This article is an elaborated version of [1], which was presented in the 5th Workshop on Planning, Perception and Navigation for Intelligent Vehicles at IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 2013.

* UMR CNRS 7253 Heudiasyc, University of Technology of Compiègne, France. E-mail: firstname.surname@hds.utc.fr

® ONERA. E-mail: firstname.surname@onera.fr

This article is organised as follows. Section II gives necessary theoretical background of the Dempster–Shafer theory of evidence. In section III, we describe the details of the proposed method, starting with the description of needed data and the purpose of each grid. Further, details on the information fusion are given. Data-dependent computation which are not in the heart of the method are described in section IV-A. Sections IV-B and IV-C present the results obtained with real-world data. Finally, section V concludes the paper and presents ideas for future work.

II. Dempster–Shafer Theory of evidence

The Dempster–Shafer Theory (DST) is a mathematical theory specially adapted to model the uncertainty and the lack of information introduced by Dempster and further developed by Shafer [3]. DST generalises the theory of probability, the theory of possibilities and the theory of fuzzy sets. In the Dempster–Shafer Theory (DST), a set $\Omega = \omega_1, \ldots, \omega_n$ of mutually exclusive propositions is called the frame of discernment (FOD). In case of closed-world hypothesis, the FOD presents also an exhaustive set. Main difference in comparison to the theory of probability is the fact that the mass of evidence is attributed not only to single hypotheses (singletons), but to any subset of the FOD, including an empty set.

As stated in the previous paragraph, beliefs about some piece of evidence are modelled by the attribution of mass to the corresponding set. This attribution of mass, called a basic belief assignment (bba), or a mass function, is defined as a mapping:

$$m(\cdot) : 2^{\Omega} \mapsto [0, 1] \tag{1}$$

$$\sum_{A \subseteq \Omega} m(A) = 1 \tag{2}$$

$$m(\emptyset) = 0 \tag{3}$$

In order to combine various information sources in the DST, there are several rules of combination. Combined mass functions have to be defined on the same FOD Ω or transform to a common frame using refining functions. A *refining* is defined as a one-to-many mapping from Ω_1 to Ω_2.

$$r : 2^{\Omega_1} \mapsto 2^{\Omega_2} \setminus \emptyset \tag{4}$$

$$r(\omega) \neq \emptyset \qquad \forall \omega \in \Omega_1 \tag{5}$$

$$\bigcup_{\omega \in \Omega_1} r(\omega) = \Omega_2 \tag{6}$$

$$r(A) = \bigcup_{\omega \in A} r(\omega) \tag{7}$$

The frame of discernment Ω_2 is then called the *refinement* of Ω_1, and Ω_1 is the *coarsening* of the Ω_2.

When combined pieces of evidence expressed by bbas are independent and both are reliable, then the conjunctive rule and Dempster's combination rule are commonly used.

In the following, let us suppose that m_1, m_2 are bbas defined on some finite frame of discernment Ω. Then, the conjunctive rule of combination denoted by $\textcircled{\cap}$ is defined as follows:

$$(m_1 \textcircled{\cap} m_2)(A) = \sum_{A=B \cap C} m_1(B) \cdot m_2(C) \tag{8}$$

The combination using the conjunctive rule can generate the mass on the empty set $m(\emptyset)$. This mass can be interpreted as the conflict measure between the combined sources. Therefore, a normalised version of conjunctive rule, called Dempster's rule and noted $\oplus$ was defined:

$$K = (m_1 \textcircled{\cap} m_2)(\emptyset) \tag{9}$$

$$(m_1 \oplus m_2)(A) = \frac{(m_1 \textcircled{\cap} m_2)(A)}{1 - K} \tag{10}$$

$$(m_1 \oplus m_2)(\emptyset) = 0 \tag{11}$$

In the DST, a discounting operation is used in order to, e.g. model information ageing. Discounting in its basic form requires to set a discounting factor α and is defined as:

$${}^{\alpha}m(A) = (1 - \alpha) \cdot m(A) \qquad \forall A \subsetneq \Omega \tag{12}$$

$${}^{\alpha}m(\Omega) = (1 - \alpha) \cdot m(\Omega) + \alpha \tag{13}$$

Decision making in DST creates sometimes a necessity of transforming a mass function into a probability function [18]. Smets and Kennes proposed so called *pignistic transformation* in [19]. Pignistic probability betP has been defined as:

$$\mathrm{betP}(B) = \sum_{A \in \Omega} m(A) \cdot \frac{|B \cap A|}{|A|} \tag{14}$$

where $|A|$ is the cardinality of the set A.

	$\emptyset$	a	b	$\Omega = \{a, b\}$
m_1	0	0.2	0.6	0.2
m_2	0	0.7	0.1	0.2
$m_1 \textcircled{\cap} m_2$	0.44	0.34	0.18	0.04
$m_1 \oplus m_2$	0	0.61	0.32	0.07
$m_1 \textcircled{\cup} m_2$	0	0.14	0.06	0.8
${}^{\alpha}m_1$	0	0.18	0.54	0.28
betP_1	0	0.3	0.7	1

TABLE I

Example of fusion rules, discounting with $\alpha = 0.1$ and pignistic probability.

Table I presents an example of different combination rules, pignistic transform and discounting operation. The Dempster–Shafer theory will serve as in further sections as the main modelling and information fusion tool in the perception system.

III. Multi-grid fusion approach

This section presents the proposed perception scheme). We use three evidential occupancy grids to model prior information, sensor acquisition and perception result. The grid construction method is described in section III-B. We detail all data processing steps in section III-D. Figure 1 presents a general functional overview of our approach. Following sections correspond to different blocks of this diagram.

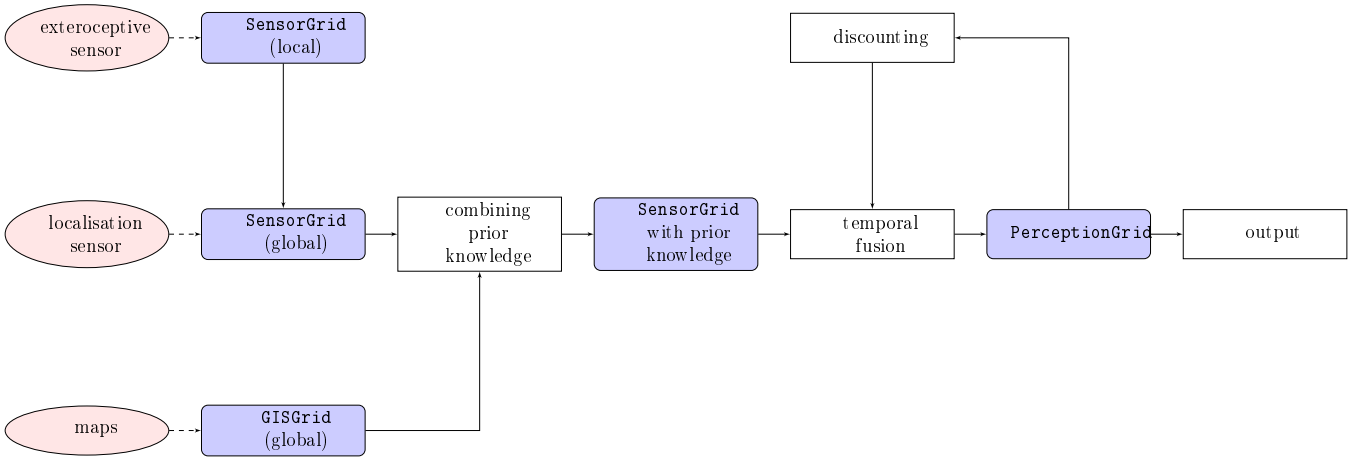


Fig. 1. Method overview.

A. Heterogeneous data sources

There are three sources in our perception system: vehicle pose, exteroceptive acquisition data and vector maps. Figure 1 illustrates all system inputs. The proposed approach is based on the hypothesis that all these information sources are available. Other hypotheses on the input data are done. Firstly, a globally referenced vehicle pose is needed to situate the system in the environment. The pose provided by a proprioceptive sensor has to be reliable and as precise as possible. It is assumed that the pose reflects closely the real state of the vehicle. Secondly, an exteroceptive sensor supplies a partial view of the environment. This sensor ought to be able to at least distinguish free and occupied space, and model it in 2D x, y or 3D x, y, z coordinates. The coordinates can be globally referenced or relative to the vehicle. A typical exteroceptive sensor capable of satisfying this assumption is a Lidar (laser range scanner), radar, or a stereo camera system. Lastly, our method tries to exploit at large the information contained in vector maps, so we assume that the maps are sufficiently rich and contain valuable accurate data. Typically, map data should contain information on the location of buildings and the model of road surface.

B. Occupancy grids

An occupancy grid models the world using a tessellated representation of spatial information. In general, it is a multi-dimensional spatial lattice with cells storing some stochastic information. In our case, each cell representing a box (a part of environment) $X \times Y$ where $X = [x_-, x_+]$, $Y = [y_-, y_+]$ stores a mass function.

1) `SensorGrid` (SG): In order to process the exteroceptive sensor data, an evidential occupancy grid is computed when a new acquisition arrives, this grid is called `SensorGrid`. Each cell of this grid stores a mass function on the FOD $\Omega_{SG} = \{F, O\}$, where:

- F free space,
- O occupied space.

The basic belief assignment reflects the sensor model.

2) `PerceptionGrid` (PG): To store the results of information fusion, an occupancy grid PG has been introduced with a FOD $\Omega_{PG} = \{N, W, I, U, S, M\}$. The choice of such a FOD is directly coupled with the objectives that we try to achieve. Respective classes represent:

- N free navigable space,
- W free non-navigable space,
- I mapped infrastructure (buildings),
- U unmapped infrastructure,
- S temporarily stopped objects,
- M mobile moving objects.

Ω_{PG} is a common frame used for scene understanding. By using PG as a cumulative information storage, we are not obliged to store preceding `SensorGrid`s.

3) `GISGrid` (GG): This grid allows the system to perform a contextual information fusion incorporating the meta-knowledge about the environment. `GISGrid` uses frame of discernment $\Omega_{GG} = \{B, R, T\}$. This FOD represents:

- B buildings,
- R roads,
- T intermediate space (e.g. pavements).

The grid can be obtained, for instance, by projection of map data, buildings and roads, onto a 2D grid with global coordinates (see Figure 5 for an example). However, the exact method of creating GG depends on available GIS information. Section IV-A.2 presents how GG is constructed.

Figure 2 demonstrates how the above grids fit together in the general scheme. It names the input sources involved at each step and shows necessary grid transformations.

C. Combining prior knowledge

In our method, prior information contained in maps serves to ameliorate the perception scheme. We have chosen to combine the prior knowledge with the sensor data of `SensorGrid`. However, the Dempster–Shafer Theory does not allow to combine sources with different frames of discernment. The frame of discernment Ω_{SG} is distinct from Ω_{GG} used in `GISGrid`. Hence, one needs to find a common frame for both sources. In order to enable the fusion of

SensorGrid (SG) and GISGrid (GG), we define refining r_{SG} to transform SensorGrid:

$$r_{SG} : 2^{\Omega_{SG}} \mapsto 2^{\Omega_{PG}} \tag{15}$$
$$r_{SG}(\{F\}) = \{N, W\} \tag{16}$$
$$r_{SG}(\{O\}) = \{I, U, S, M\} \tag{17}$$
$$r_{SG}(A) = \bigcup_{\theta \in A} r_{SG}(\theta) \tag{18}$$

and refining r_{GG} to transform GISGrid:

$$r_{GG} : 2^{\Omega_{GG}} \mapsto 2^{\Omega_{PG}} \tag{19}$$
$$r_{GG}(\{B\}) = \{I\} \tag{20}$$
$$r_{GG}(\{R\}) = \{N, W, S, M\} \tag{21}$$
$$r_{GG}(\{T\}) = \{N, W, S, M, U\} \tag{22}$$
$$r_{GG}(A) = \bigcup_{\theta \in A} r_{GG}(\theta) \tag{23}$$

Above defined refinings allows us to combine prior knowledge included in GISGrid with instantaneous grid obtained from sensor(s). The refined mass function can be expressed as:

$$m_{SG}^{\Omega_{PG}}(r_{SG}(A)) = m_{SG}^{\Omega_{SG}}(A) \qquad \forall A \subseteq \Omega_{SG} \tag{24}$$
$$m_{GG}^{\Omega_{PG}}(r_{GG}(A)) = m_{GG}^{\Omega_{GG}}(A) \qquad \forall A \subseteq \Omega_{GG} \tag{25}$$

Then, Dempster's rule described in section II is applied to each cell in order to exploit the prior information included in GG:

$$m'^{\Omega_{PG}}_{SG,t} = m_{SG,t}^{\Omega_{PG}} \oplus m_{GG}^{\Omega_{PG}} \tag{26}$$

We have chosen to use the Dempster's rule of combination, since the GIS data and the sensor data are independent. Besides, we suppose that both sources are reliable, even if errors are possible. In the end of this stage, we obtain a grid being a combination of the sensor data, SensorGrid, with the prior knowledge from GISGrid.

D. Temporal fusion

The role of the fusion operation is to combine current sensor acquisition with preceding perception result. The sensor acquisition input is already combined with prior information as described in preceding paragraphs. We propose to exploit dynamic characteristics of the scene by analysing produced conflict masses. As the preceding perception result PerceptionGrid is partially out-of-date at the moment of fusion, the contextual discounting operation is employed to model this phenomena [20]. Moreover, an accumulator of occupancy is introduced and a mass function specialisation is performed to distinguish mobile, but temporarily stopped objects.

1) Computing conflict masses: To distinguish between two types of conflict which arise from the fact that the environment is dynamic, the idea from [21] is used. $\emptyset_{FO}$ denotes the conflict induced when a free cell in PG is fused with an occupied cell in SG. Similarly, $\emptyset_{OF}$ indicates the conflicted mass caused by an occupied cell in PG fused with a free cell in SG.

Conflict masses are calculated using the formulas:

$$m_{PG,t}(\emptyset_{OF}) = m_{PG,t-1}(O) \cdot m_{SG,t}(F) \tag{27}$$
$$m_{PG,t}(\emptyset_{FO}) = m_{PG,t-1}(F) \cdot m_{SG,t}(O) \tag{28}$$

where $m(O) = \sum_{A \subseteq \{I,U,S,M\}} m(A)$ and $m(F) = \sum_{A \subseteq \{N,W\}} m(A)$. In an error-free case, these conflicts represent, respectively, the disappearance and the appearance of an object in a given cell.

2) PerceptionGrid specialisation using an accumulator: Mobile object detection is an important issue in dynamic environments. We propose the introduction of an accumulator ζ in each cell in order to include temporal information on the cell occupancy. For this purpose, incrementation and decrementation steps $\delta_{inc} \in [0,1]$, $\delta_{dec} \in [0,1]$, as well as threshold values γ_O, $\gamma_\emptyset$ have been defined.

$$\zeta^{(t)} = \min\left(1, \zeta^{(t-1)} + \delta_{inc}\right) \tag{29}$$
if $m_{PG}(O) \geq \gamma_O$
and $m_{PG}(\emptyset_{FO}) + m_{PG}(\emptyset_{OF}) \leq \gamma_\emptyset$
$$\zeta^{(t)} = \max\left(0, \zeta^{(t-1)} - \delta_{dec}\right) \tag{30}$$
if $m_{PG}(\emptyset_{FO}) + m_{PG}(\emptyset_{OF}) > \gamma_\emptyset$
$$\zeta^{(t)} = \zeta^{(t-1)} \tag{31}$$
otherwise (32)

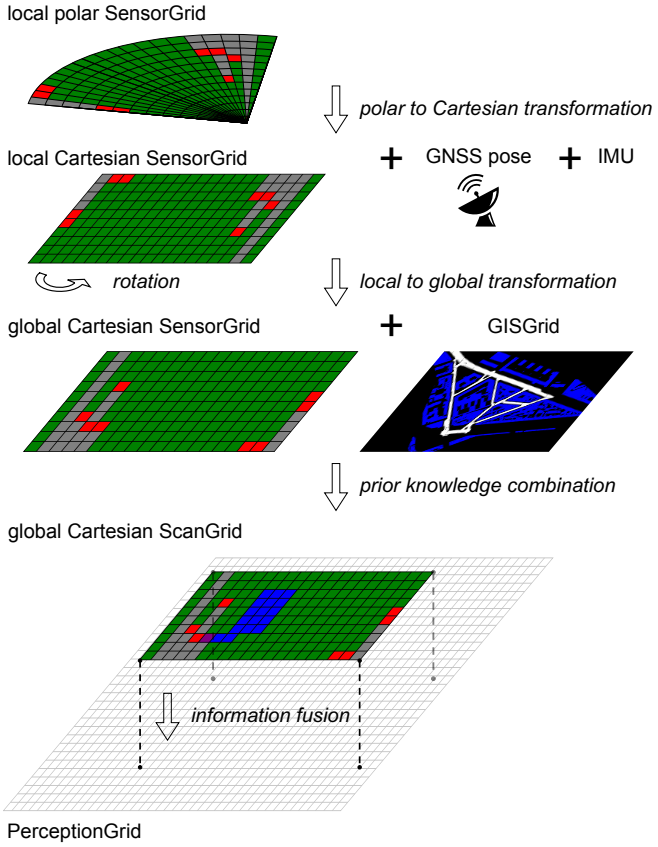


Fig. 2. Illustration of grid transformations.

Using ζ values, we impose a specialisation of mass functions in PG using the equation:

$$m'_{PG,\,t}\,(A) = S(A,B) \cdot m_{PG,\,t}(B) \tag{33}$$

where specialisation matrix $S(\cdot,\cdot)$ is defined as:

$$\begin{aligned} S(A\backslash\{M\},\,A) &= \zeta && \forall A \subseteq \Omega_{PG} \text{ and } \{M\} \in A \\ S(A,\,A) &= 1-\zeta && \forall A \subseteq \Omega_{PG} \text{ and } \{M\} \in A \\ S(A,\,A) &= 1 && \forall A \subseteq \Omega_{PG} \text{ and } \{M\} \notin A \\ S(\cdot,\,\cdot) &= 0 && \text{otherwise} \end{aligned} \tag{34}$$

The idea behind the specialisation matrix and the accumulator is that the mass attributed to set N, S, M or S, M is transferred to set N, S or S, respectively. The transferred mass value is proportional to the time that the cell stayed occupied. In this way, moving objects are differentiated from static or stopped objects.

3) Fusion rule: An important part of the method consists in performing the fusion operation of a discounted and specialized `PerceptionGrid` from preceding epoch ${}^{\alpha}m'_{PG,\,t-1}$ with a SG combined with prior knowledge from current epoch $m'_{SG,\,t}$. The discounting operation is presented in section II and the specialisation is described in the preceding paragraph. In the section III-C, combination of prior knowledge with `SensorGrid` is demonstrated.

$$m_{PG,\,t} = {}^{\alpha}m'_{PG,\,t-1} \circledast m'_{SG,\,t} \tag{35}$$

The fusion rule $\circledast$ is a modified conjunctive rule adapted to mobile object detection. There are of course many different rules that could be used, but in order to distinguish between moving and stationary objects some modifications had to be performed. These modifications consist in transferring the mass corresponding to a newly appeared object $\emptyset_{FO}$ to the class of moving objects M as described by the equation 36. Please remind that $\textcircled{\cap}$ denotes the conjunctive fusion rule.

$$\begin{aligned} (m_1 \circledast m_2)\,(A) &= (m_1 \textcircled{\cap} m_2)\,(A) \\ &\qquad \forall A \subsetneq \Omega \wedge A \neq M \\ (m_1 \circledast m_2)\,(M) &= (m_1 \textcircled{\cap} m_2)\,(M) + (m_1 \textcircled{\cap} m_2)\,(\emptyset_{FO}) \\ (m_1 \circledast m_2)\,(\Omega) &= (m_1 \textcircled{\cap} m_2)\,(\Omega) + (m_1 \textcircled{\cap} m_2)\,(\emptyset_{OF}) \\ (m_1 \circledast m_2)\,(\emptyset_{FO}) &= 0 \\ (m_1 \circledast m_2)\,(\emptyset_{OF}) &= 0 \end{aligned} \tag{36}$$

All the above steps allow the construction of a PG containing reach information on the environment state, including the knowledge on mobile and static objects.

E. Fusion rule behaviour

Proposed fusion scheme behaves differently depending on the context. In this section, we describe briefly the behaviour of the fusion rule. For an in-depth analysis, the reader is invited to read [22]. *Context* stands for prior knowledge information contained in `GISGrid`. To demonstrate the effect of the fusion operator, we have chosen two particular cases, which clearly represent different contexts.

Building context: In the building context, the fusion rule behaves as Yager's rule, which consists in transferring the conflict mass to unknown class Ω [23]. This behaviour is relevant, since it is assumed that no mobile obstacles are present in this context. Therefore, only free space and infrastructure are to be distinguished.

Road and intermediate space: The conflict management adapted to the perception scheme directs mass attribution to moving obstacles (class M). The introduction of occupied space counter and `PerceptionGrid` specialisation (see section III-D.2) permits to transfer a part of the mass from "moving or other" class to "other", where other is context-dependent.

F. Illustrative examples

This section aims to present the behaviour of our perception system on a pedagogical example. The example is composed of three 1D-grids (for simplification, here) evolving in time, and so the scenario has to be read from top to bottom, line by line. Colours are determined by the mass function of the cell according to the legend. Please note that a mass function can contain more than one focal element so the effective colour is a mix of corresponding classes.

In the scene (see Figure 3), there are 3 pedestrians: one crossing the road, one staying on the road and one staying on the pavement. `GISGrid` represents our prior knowledge obtained from the digital map. Both sides are known to be pavement (intermediate space T) and the middle is supposed to be the road surface R. `SensorGrid` is a representation of current sensor data. The sensor provides in this situation only free/occupied information. `PerceptionGrid` is the step-by-step result of the perception system. Its initial state is a complete ignorance (all mass attributed to unknown Ω). As the moving pedestrian walks, the movement is detected in different grid cells. Behind the pedestrian, previously occupied cells become gradually free. Free, navigable and non-navigable, cells have their masses increased as the sensor confirms the same information. The information relative to the two stopped pedestrians is processed differently thanks to the map information. The one on the road is considered as a movable MS object at first, since initially the cell was unknown. Then, as time evolves, it is considered as stopped, which illustrates how the accumulator and the specialisation work. This change is visible as the colour changes from purple to blue. The other pedestrians on the pavement are seen as moving or stopped or unmapped obstacle MSU at the beginning. At the end, the one staying on the pavement is detected as a stopped obstacle *or* as unmapped infrastructure SU. Indeed, it is impossible to distinguish between a stopped pedestrian and, for instance, a light post. The pedestrian crossing the road is quickly detected as moving M.

IV. Experimental results

A. Experimental setup

1) Dataset: The data set used for experiments was acquired in the 12th district of Paris using our vehicular platform Carmen (see figure 4). The overall length of the

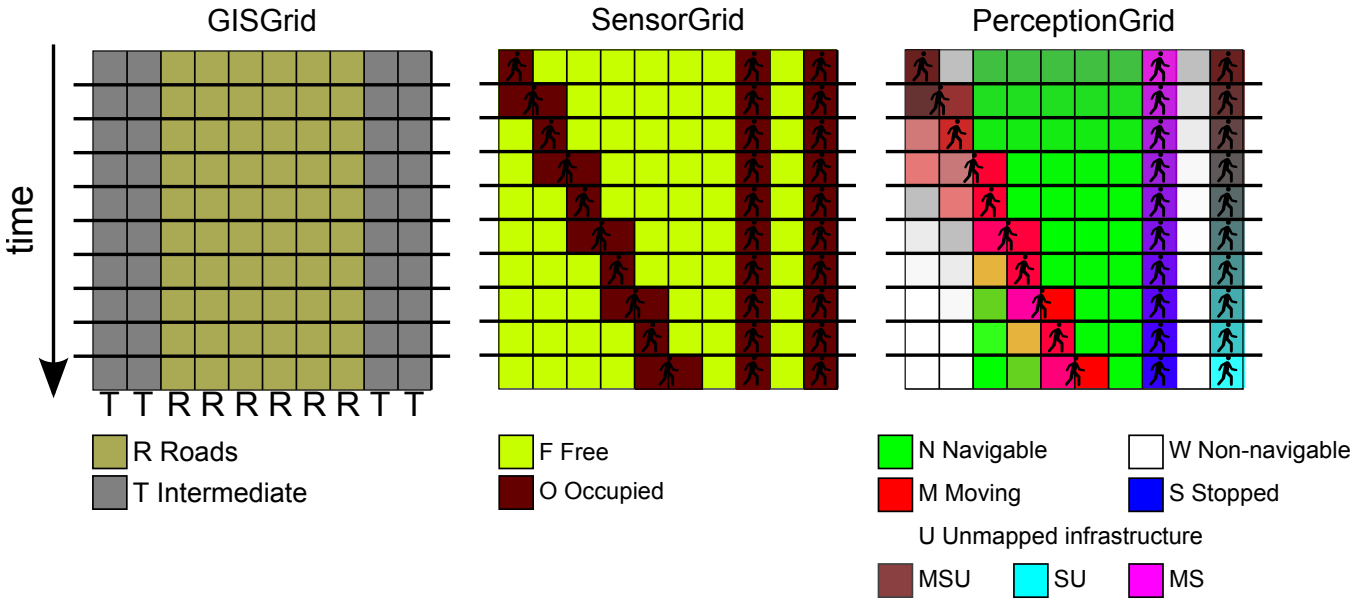


Fig. 3. Example with three pedestrians. T represents the intermediate space and R – road surface.

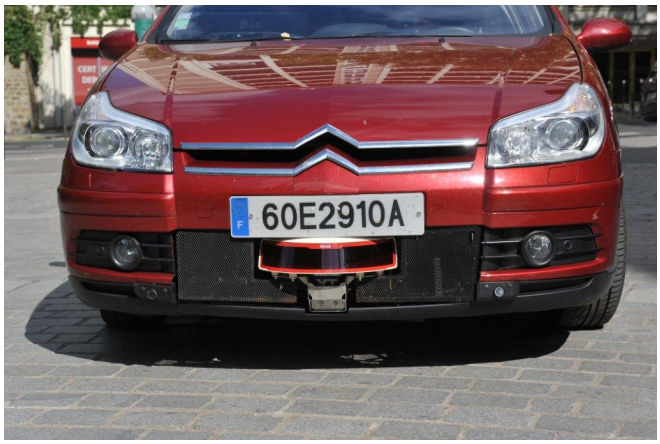


Fig. 4. Vehicle Carmen with Alaska XT lidar embedded in the front bumper.

trajectory was about 9 kilometres. The vehicle pose comes from a system based on a NovAtel SPAN-CPT inertial measurement unit (IMU). The system provides precise positioning with high confidence. Our main source of information about the environment is an IBEO Alaska XT lidar able to provide a cloud of about 800 points 10 times per second. The digital maps that we use were provided by the French National Geographic Institute (IGN) and contain 3D building models as well as the road surface. We also performed successful tests with freely available *OpenStreetMap* project 2D maps [12], but here we limited the use to building data. We assume the maps to be accurate and up-to-date.

2) `GISGrid` construction: The map data can be represented by two sets of polygons defining the 2D position of buildings and road surface by, respectively,

$$\mathcal{B} = \left\{ b_i = \begin{bmatrix} x_1 x_2 \dots x_{m_i} \\ y_1 y_2 \dots y_{m_i} \end{bmatrix}, i \in [0, n_B] \right\} \tag{37}$$

$$\mathcal{R} = \left\{ r_i = \begin{bmatrix} x_1 x_2 \dots x_{m_i} \\ y_1 y_2 \dots y_{m_i} \end{bmatrix}, i \in [0, n_R] \right\} \tag{38}$$

Our dataset satisfies the condition: $\mathcal{B} \cap \mathcal{R} = \emptyset$. An example of `GISGrid` is presented in Figure 5 juxtaposed with an aerial photo with 3D building models. Classes B, R, T characterise the meta-information inferred from geographic maps. Set A denotes then all other strict subsets of Ω. For instance, on the road surface R, we encourage the existence of free navigable space N as well as stopped S and moving M objects. Analogically, building information B fosters mass transfer to I. Lastly, T denotes the intermediate area, e.g. pavements, where mobile and stationary objects as well as small urban infrastructure can be present. T can contain free non-navigable space W as well. Please note that neither buildings nor roads are present, so the existence of mapped infrastructure I can be excluded, but the presence of the other classes cannot. Also, a level of confidence β is defined for each map source, possibly different for each context (in our case: $\beta_B = \beta_R = \beta_T = \beta$). Let $m_{GG}\{X, Y\}$ be the mass function in the cell at coordinates $\{X, Y\} = \{[x_-, x_+], [y_-, y_+]\}$ and $\tilde{x} = \frac{x_- + x_+}{2}$, $\tilde{y} = \frac{y_- + y_+}{2}$, then:

$$m_{GG}\{X, Y\}(B) = \begin{cases} \beta_B & \text{if } (\tilde{x}, \tilde{y}) \in b_i \\ 0 & \text{otherwise} \end{cases} \tag{39}$$
$$\forall i \in [0, n_B]$$

$$m_{GG}\{X, Y\}(R) = \begin{cases} \beta_R & \text{if } (\tilde{x}, \tilde{y}) \in r_i \\ 0 & \text{otherwise} \end{cases} \tag{40}$$
$$\forall i \in [0, n_R]$$

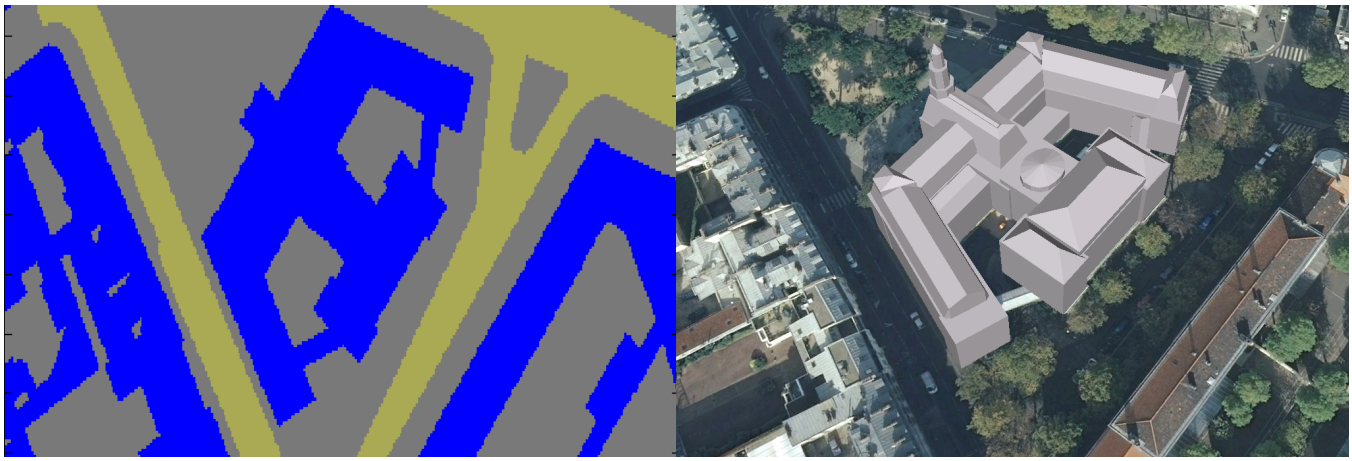


Fig. 5. Paris 12th district cityhall. Left: `GISGrid` constructed using our dataset. Right: 3D view of the area, source: Google Earth. Buildings in blue, roads in khaki, intermediate space in gray.

$$m_{GG}\{X,Y\}(T) = \begin{cases} 0 & \text{if } (\tilde{x},\tilde{y}) \in b_i \vee (\tilde{x},\tilde{y}) \in r_j \\ \beta_T & \text{otherwise} \end{cases} \tag{41}$$

$$\forall i \in [0, n_B], \forall j \in [0, n_R]$$

$$m_{GG}\{X,Y\}(\Omega) = \begin{cases} 1-\beta_B & \text{if } (\tilde{x},\tilde{y}) \in b_i \\ 1-\beta_R & \text{if } (\tilde{x},\tilde{y}) \in r_i \\ 1-\beta_T & \text{otherwise} \end{cases} \tag{42}$$

$$\forall i \in [0, n_B], \forall j \in [0, n_R]$$

$$m_{GG}\{X,Y\}(A) = 0 \tag{43}$$

$$\forall A \subsetneq \Omega \text{ and } A \notin \{B, R, T\}$$

3) Sensor model: This section describes the way in which the data obtained from the lidar are transformed into `SensorGrid`. If another exteroceptive sensor is used, one has to define an appropriate model. The model used in the presented method is based on the one described in [14] which takes into account the measurement uncertainties and occlusions. As demonstrated in Figure 2, the sensor provides a polar grid. This grid undergoes several transformation before being fused into `PerceptionGrid`. Our lidar model uses frame of discernment Ω_{SG} with two classes: free F and occupied O. Each scan point (lidar impact) results in a grid cell being set with high value of occupied mass. All the cells before the first cell in each angular sector have their main part of the mass attributed to free. Small portion of the mass is set to Ω in order to take into account sensor resolution and grid discretisation errors. The state of other cells, those between obstacles and behind obstacles, cannot be determined and so all the mass goes to the unknown Ω.

4) Parameters: The size of the grid cell in the occupancy grids was set to 0.5 m, which is sufficient to model a complex environment with mobile objects. We have defined the map confidence factor β by ourselves, but ideally, it should be given by the map provider. β describes data currentness (age), errors introduced by geometry simplification and spatial discretisation. β can also be used to depict the localisation accuracy. In our case, $\beta = 0.005$, since we suppose our maps precise and accurate. Other parameters, such as counter steps δ_{inc}, δ_{dec} and thresholds γ_O, $\gamma_\emptyset$ used for mobile object detection determine the sensitiveness of mobile object detection and were set by manual tuning.

B. Results of obstacle detection

The results for a particular instant of the approach tested on real data are presented on Figures 6 and 7. The reported scenes were recorded while the vehicle was moving to illustrate the performance in real urban traffic conditions, typically at the speed of 40 km/h.

Topmost images show camera captures, middle images present `PerceptionGrid` in a fixed Cartesian frame, zoomed in around the vehicle location. The visualisation of PG has been obtained by attributing to each class a colour proportional to the pignistic probability betP and calculating the mean colour. Images containing grids contain some markers to show the vehicle position (small red cross) and vehicle speed vector (black arrow). Light dashed white lines show the approximate camera's field of view limits in order to link the image with the grid. Please notice that the field of view of the lidar is wider than the one of the camera and is not shown for clarity. Bottommost images reflect the result of a decision rule done by thresholding pignistic probabilities (see Equation 14). The different thresholds were set to 0.5 except for class S for which the threshold was 0.35, since we wanted to magnify the effect of detection of obstacles stopped for a short moment.

Figure 6 presents quite a complex scene with multiple moving vehicles and a few stopped ones. The two moving motorcycles and the moving car on the opposite lane are clearly detected as shown by red cells on the bottom figures. Behind these moving cells, the state of the space is unknown, which is consistent with the lidar capabilities. The car in front, waiting at a traffic light, has been detected as stopped (see the blue cells right in the direction of the arrow).

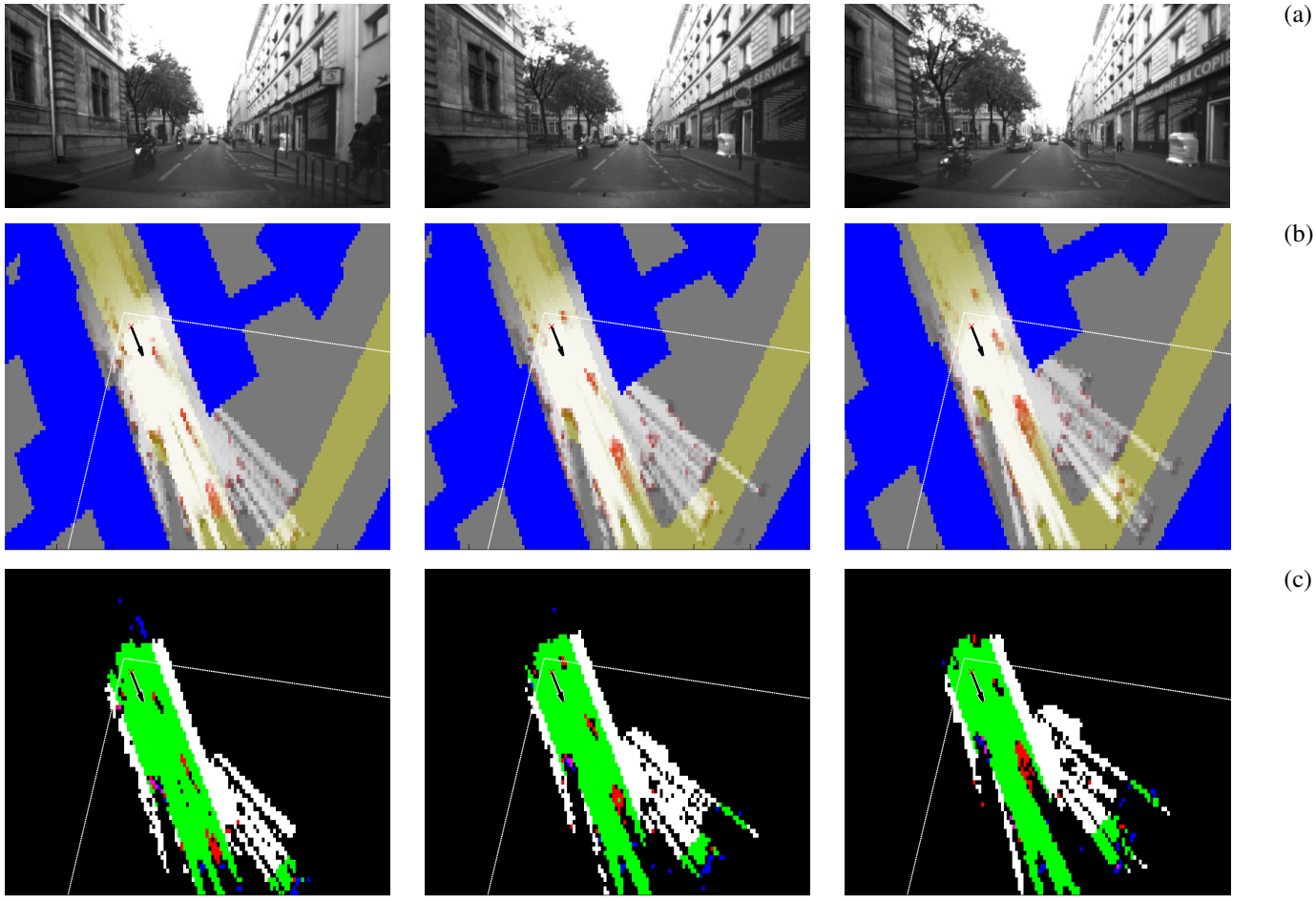


Fig. 6. Scene 1 evolution: time goes from left to right. From top to bottom: (a) scene capture, (b) `PerceptionGrid` pignistic probability, (c) simple decision rule to detect free space, moving and stopped obstacles, Colour code for figures (3): green – navigable free space, white – non-navigable free space, red – moving objects, blue – stopped objects, black – unknown. Infrastructure (classes U and I) has not been visualised in the bottom figures.

Similarly, cars parked on the left side road are detected as stopped (blue cells in the bottom right of the grids). One can notice that even if these vehicles are hardly visible on the camera images, they have been detected by the perception system. When the size of the objects become small, the lidar can miss them. This can induce some artefacts in the perception scheme making for instance posts of traffic signs oscillating between moving and stopped. This explains the isolated red/blue cells in the grid.

Figure 7 presents another complex scene containing three cars moving in the opposite direction (visible only in some photos), one parked car, one parked bus and a motorcycle going in the same direction as the equipped vehicle. Moving cars (in red) are well-distinguished in the bottom images. Navigable (green) and non-navigable (white) spaces are well characterised and clearly separated. The partially visible bus and the car parked on the left (blue) are well detected as well.

The additional information provided by the map clearly enhances the driving scene understanding. The system is able to make a clear difference between moving (red cells) and stopped (blue) objects. Also, we have noticed on other sequences that stopped objects are perceived distinctly from infrastructure when prior map information is available. In addition, thanks to the prior knowledge, stationary objects such as infrastructure are distinguished from stopped objects on the road. This is a behaviour similar to one of Capelle et al. system that uses 3D city models [24].

Finally, the effect of discounting is noticeable particularly behind the vehicle, as the information about the environment is being forgotten with different rates thanks to the map. The grid cells get discounted, so the mass on the different classes diminishes gradually. The thresholded plots show that the stopped information is more remanent as some blue cells are left behind.

C. Results of free space detection

Figure 8 is the result of cumulating subsequent grids for navigable and non-navigable free spaces after having executed the pignistic decision rule. These figures show all the cells that were detected as free either on the road or on the pavement at least at one moment of the test sequence. These cells are shown in green for navigable space and in white for non-navigable space. In violet, we superposed the prior map information about road surface from `GISGrid`. In

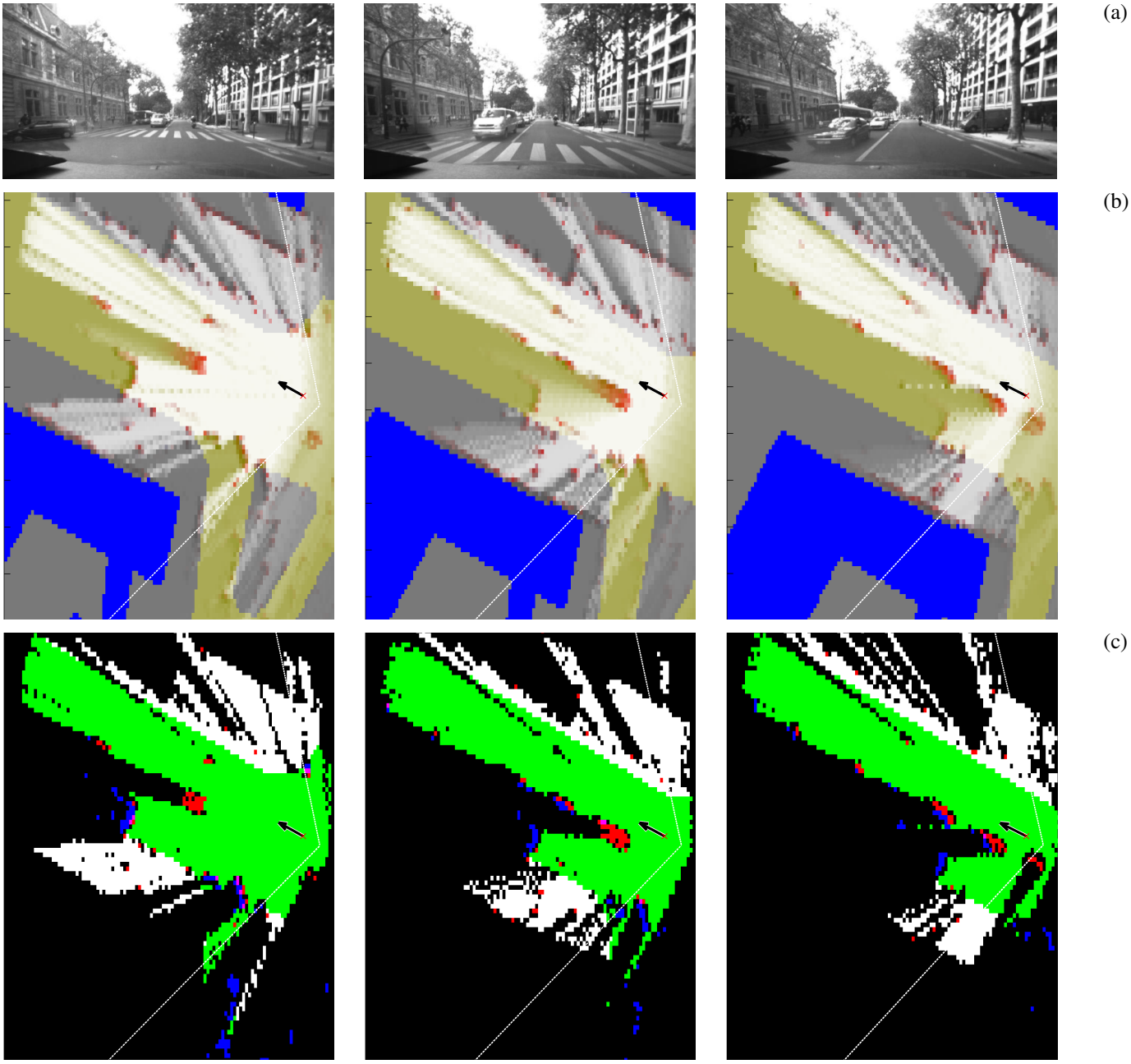


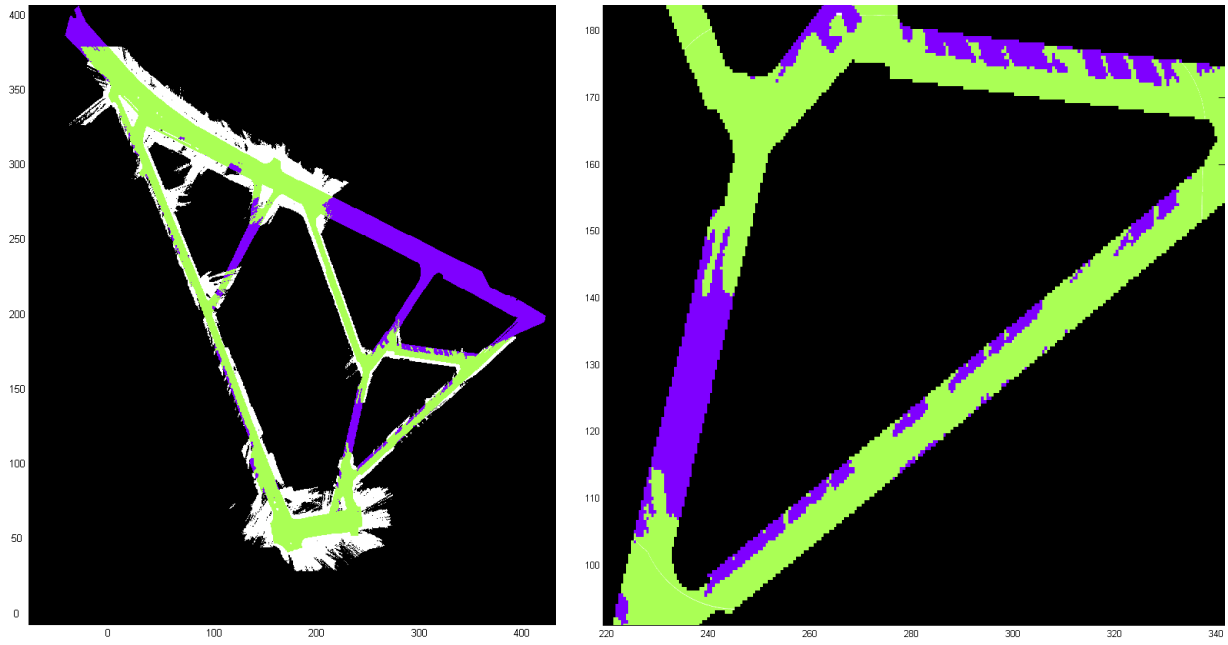
Fig. 7. Scene 2 evolution. Same display as Figure 6.

such a way, one can identify areas that the vehicle perceived during the test. More interestingly, one can clearly recognise on Figure 8(b) the places where cars or other vehicles were parked. On the other hand, the non-navigable free space on Figure 8(a) (in white) exhibits zones that are normally useless for navigation but that could be useful for safety manoeuvres like the avoidance of a pedestrian collision.

V. Conclusion and perspectives

A new mobile perception scheme based on prior map knowledge has been introduced. Geographic information is exploited to reduce the number of possible hypotheses delivered by an exteroceptive source. A modified fusion rule taking into account the existence of mobile objects has been defined. Furthermore, the variation in information lifetime has been modelled by the introduction of contextual discounting.

In the future, we anticipate removing the hypothesis that the map is accurate. This approach will entail considerable work on creating appropriate error models for the data source. This will be a step towards the use of our approach in autonomous navigation. Another perspective is the use of reference data to validate the results, choose the most appropriate fusion rule and learn algorithm parameters. We envision using map information to predict object movements. It rests also a future work to exploit fully the 3D map



(a) Whole test sequence area.

(b) Zoom on navigable space N.

Fig. 8. Free spaces N (in green) and non-navigable (in white) cumulated over the complete test sequence. Road surface from maps in violet. Both axes in meters.

information.

ACKNOWLEDGEMENTS

This work was supported by the French Ministry of Defence DGA (Direction Générale de l'Armement), with a Ph.D. grant delivered to Marek Kurdej.

REFERENCES

[1] M. Kurdej, J. Moras, V. Cherfaoui, and P. Bonnifait, "Enhancing Mobile Object Classification Using Geo-referenced Maps and Evidential Grids", in *IEEE/RSJ International Conference on Intelligent Robots and Systems. 5th Workshop on Planning, Perception and Navigation for Intelligent Vehicles*, Tokyo, Nov. 2013.

[2] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. Cambridge, Massachusetts, USA: MIT Press, 2005.

[3] G. R. Shafer, *A Mathematical Theory of Evidence*, en. Princeton University Press, 1976, pp. 1–297.

[4] A. Elfes, "Using Occupancy Grids for Mobile Robot Perception and Navigation", en, *Computer Journal*, vol. 22, no. 6, pp. 46–57, Jun. 1989.

[5] D. Pagac, E. M. Nebot, and H. Durrant-Whyte, "An evidential approach to map-building for autonomous vehicles", *IEEE Transactions on Robotics and Automation*, vol. 14, no. 4, pp. 623–629, 1998.

[6] S. M. Oh, S. Tariq, B. N. Walker, and F. Dellaert, "Map-based Priors for Localization", en, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sendai: IEEE, 2004, pp. 2179–2184.

[7] C. Fouque, P. Bonnifait, and D. Bétaille, "Enhancement of Global Vehicle Localization using Navigable Road Maps and Dead-Reckoning", en, in *Position, Location and Navigation Symposium*, Monterey, CA: IEEE, May 2008, pp. 1286–1291.

[8] F. Moosmann, "Interlacing Self-Localization, Moving Object Tracking and Mapping for 3D Range Sensors", PhD thesis, Karlsruher Institut für Technologie (KIT), 2012.

[9] J. Moras, S. A. Rodriguez Flores, V. Drevelle, G. Dherbomez, V. Cherfaoui, and P. Bonnifait, "Drivable Space Characterization Using Automotive Lidar and Georeferenced Map Information", in *IEEE Intelligent Vehicles Symposium*, Alcalá de Henares, 2012, pp. 778–783.

[10] J. Wei, J. M. Snider, J. Kim, J. M. Dolan, R. Rajkumar, and B. Litkouhi, "Towards a Viable Autonomous Driving Research Platform", en, in *IEEE Intelligent Vehicles Symposium*, Gold Coast, Jun. 2013, pp. 763–770.

[11] K. Irie and M. Tomono, "Road Recognition from a Single Image using Prior Information", en, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Tokyo, Nov. 2013, pp. 1938–1945.

[12] *OpenStreetMap*, 2013. [Online]. Available: `http://www.openstreetmap.org`.

[13] W. Xiao, S. Xu, S. Oude Elberink, and G. Vosselman, "Change detection of trees in urban areas using multi-temporal airborne lidar point clouds", en, in *Remote Sensing of the Ocean, Sea Ice, Coastal Waters, and Large Water Regions 2012. Proceedings of the SPIE*, C. R. Bostater, S. P. Mertikas, X. Neyt, C. Nichol, D. Cowley, and J.-P. Bruyant, Eds., vol. 8532, Oct. 2012.

[14] J. Moras, V. Cherfaoui, and P. Bonnifait, "Credibilist Occupancy Grids for Vehicle Perception in Dynamic Environments", en, in *IEEE International Conference on Robotics and Automation*, Shanghai, May 2011, pp. 84–89.

[15] K. M. Wurm, A. Hornung, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: A Probabilistic, Flexible, and Compact 3D Map Representation for Robotic Systems", en, in *IEEE International Conference on Robotics and Automation Workshop*, May 2010. [Online]. Available: `http://octomap.sourceforge.net/`.

[16] M. Dawood, C. Cappelle, M. E. E. Najjar, M. Khalil, and D. Pomorski, "Vehicle geo-localization based on IMM-UKF data fusion using a GPS receiver, a video camera and a 3D city model", in *IEEE Intelligent Vehicles Symposium*, 2011, pp. 510–515.

[17] M. Hentschel and B. Wagner, "Autonomous Robot Navigation Based on OpenStreetMap Geodata", in *International IEEE Annual Conference on Intelligent Transportation Systems*, Madeira Island, Sep. 2010, pp. 1645–1650.

[18] P. Smets, "Decision Making in the TBM: the Necessity of the Pignistic Transformation", *International Journal of Approximate Reasoning*, vol. 38, no. 2, pp. 133–147, 2005.

[19] ——, “What is Dempster-Shafer’s model?”, in *Advances in the Dempster-Shafer Theory of Evidence*, R. R. Yager, M. Fedrizzi, and J. Kacprzyk, Eds. John Wiley & Sons, 1994, pp. 5–34.

[20] M. Kurdej and V. Cherfaoui, “Conservative, Proportional and Optimistic Contextual Discounting in the Belief Functions Theory”, en, in *16th International Conference on Information Fusion*, Istanbul, Jul. 2013.

[21] J. Moras, V. Cherfaoui, and P. Bonnifait, “Moving Objects Detection by Conflict Analysis in Evidential Grids”, en, *IEEE Intelligent Vehicles Symposium*, pp. 1120–1125, Jun. 2011.

[22] M. Kurdej, J. Moras, V. Cherfaoui, and P. Bonnifait, “Controlling Remanence in Evidential Grids Using Geodata for Dynamic Scene Perception”, en, *International Journal of Approximate Reasoning*, vol. 55, no. 1, pp. 355–375, Mar. 2014.

[23] R. R. Yager, “On the Dempster-Shafer framework and new combination rules”, en, *International Journal of Information Sciences*, vol. 41, no. 2, pp. 93–137, Mar. 1987.

[24] C. Cappelle, M. E. El Najjar, F. Charpillet, and D. Pomorski, “Virtual 3D City Model for Navigation in Urban Areas”, en, *Journal of Intelligent Robotic Systems*, vol. 66, no. 3, pp. 377–399, 2012.