An evidential filter for indoor navigation of a mobile robot in dynamic environment

Quentin Labourey, Olivier Aycard, Denis Pellerin, Michele Rombaut, and
Catherine Garbay

Univ. Grenoble Alpes

Abstract. Robots are destined to live with humans and perform tasks
for them. In order to do that, an adapted representation of the world
including human detection is required. Evidential grids enable the robot
to handle partial information and ignorance, which can be useful in var-
ious situations. This paper deals with an audiovisual perception scheme
of a robot in indoor environment (apartment, house..). As the robot
moves, it must take into account its environment and the humans in
presence. This article presents the key-stages of the multimodal fusion:
an evidential grid is built from each modality using a modified Dempster
combination, and a temporal fusion is made using an evidential filter
based on an adapted version of the generalized bayesian theorem. This
enables the robot to keep track of the state of its environment. A deci-
sion can then be made on the next move of the robot depending on the
robot's mission and the extracted information. The system is tested on
a simulated environment under realistic conditions.

Keywords: Active multimodal perception, Evidential filtering, Mobile
robot

1 Introduction

Perceptually-driven robots have raised increased interest in different domains
(autonomous vehicles [7], medecine [1], social robotics [10][4]). More precisely,
the notion of companion robots, i.e. robots that are able to live with and per-
form services for humans has appeared in the litterature. Those robots need to
be autonomous: they must monitor and estimate the needs of the persons in
their environment and react accordingly by positioning themselves in a socially
acceptable way. In order to do that, the humans in the scene must be detected
and tracked in time, in an adequate world representation. It is in this context
that this work takes place. In this paper, we consider the case of a small indoor
mobile robot, equipped with video and depth (RGB-D) and audio sensors, that
must keep track of humans in its surrounding in order to accomplish a mission.
Each sensor is able to provide information on the possible presence of humans
at different locations with respect to the robot, with a certain degree of uncer-
tainty. The robot must then fuse at best the information at hand, while taking
into account this uncertainty, in order to localize humans properly.

This requires the robot to be able to create and update a map of its environment and to localize itself in this map. The most common kind of map is the grid, where each cell of the grid represents a spatial portion of the world and contains information as to the content of this portion. Until recently, grids contained only information of occupancy to enable the robot to move around obstacles and detect moving objects [9] [2].

In our case, we want the grid to contain more information in order to extract relevant information for the mission, i.e. human positions. Moreover uncertainty and doubt are crucial to the problem, as the perception algorithms and sensors are not perfect, and the spatial field of detection is only partial: if the robot suspects the presence of a person at a position in space but with a high level of doubt, it can be part of its mission to move and position itself in order to confirm or infirm the presence of a person at that position. In the same way a spatial zone that was never perceived before (ignorance about the content) can be of high interest for the robot, as there might be humans in that zone. A framework adapted to this representation of knowledge is the evidential framework [8]. Interesting works already exist on evidential grids [5][3], mainly in the field of autonomous cars: in both works, a fusion scheme is presented that takes sensor data (and a priori knowledge in the case of [3]) and fuse it into a time-evolutive model in order to produce an enhanced grid containing information relevant to vehicle navigation. However in both cases the only exteroceptive sensor used is the LIDAR, which means that no multimodality is used. Moreover the perception is only used to increase the accuracy of navigation and does not take into account the mission of the robot.

We propose a fusion scheme based on extracted data from the sensors at each observation time. The information extracted by the sensors are directly integrated into evidential sensor grids that are fused together to produce a perception grid containing the information accross all sensors. To increase the robustness and add information about dynamicity, this grid is fused into an evolutive evidential model. This enables the robot to keep all the extracted data directly into an information grid from which decisions concerning its next motion can be taken depending on its mission.

2   World model and evidential grids

A classic way for a robot to model its environment is the occupancy grid. An occupancy grid is a grid in which each cell represents a spatial portion of the robot's environment. Inside each cell is an information about the state of occupancy of this cell. The most common type of occupancy grid is an estimation for each cell of a probability distribution over the two states "occupied" or "free".

We consider the case of a small indoor mobile robot, equipped with a kinect-type sensor (video and depth information), as well as a couple of microphones (audio information). This robot has missions to fulfill in dynamic environment, in human presence (e.g. monitor the state of the persons in the environment, patrol and search for humans in order to transmit their position, etc...). In order to accomplish its mission, the robot must navigate through its environment, i.e.

know where it is allowed to go without collisions with its environment and what pathes to consider to reach a goal, and at the same time extract information relevant to its mission, i.e. places where the humans might be present, places where objects might be, and the dynamics of the scene (mobility, staticity of its surrounding). The robot must also keep in memory the places where it does not know the content of the scene, as they might become the next important locations to visit.

We are thus trying to estimate the state of the spatial surrounding of the robot, in the state space $\Omega = \{H, O, F\}$, where:

- H: Presence of a human, mobile or static
- O: Presence of a non-human object, mobile or static
- F: Free navigable space

As the perceptual capabilities of the robot are not perfect (faulty sensors or perception algorithms), a deterministic approach can present a quantity of drawbacks. To have all the information relevant to navigation and the mission of the robot in one unified stochastic representation, we propose a fusion method based on evidential grids [3].

An *evidential* grid is a grid in which each cell contains information about the state of a portion of the robot's environment in the form of a belief function. The outline of the method is presented in figure 1:
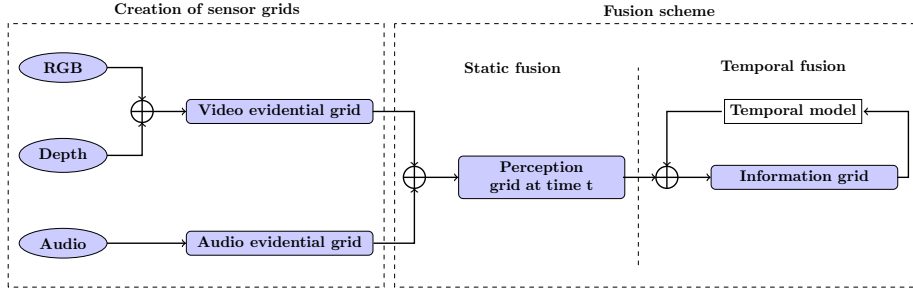

Fig. 1: Global fusion scheme

At each time step, information is extracted from sensors, and sensor grids are created accordingly. The creation of such grids is detailed in section 3. The sensor grids are then fused together in order to obtain a perception grid. This perception grid is then fused in an evolutive fusion model, in order to extract information about the dynamics of the scene. The creation of the perception grid and the temporal fusion are detailed in section 4.

Throughout the article the following notations are used: m^Ω, q^Ω, and bel^Ω represent respectively a basic belief assignment (bba), commonality function, and belief function over discernment space Ω.

$m^{\Omega}[S], q^{\Omega}[S]$, and $bel^{\Omega}[S]$ represent respectively a conditional basic belief assignment, commonality function, and belief function over discernment space Ω, which represents the belief repartition knowing hypothesis H to be true.

3   Robot perceptual system and sensor grids
3.1   Perceptual capabilities of the robot

It is assumed that an occupancy grid of the environment has been built in a previous patrol of the robot, thanks to a Simultaneous Localization and Mapping (SLAM) algorithm, and contain information about purely static objets (wall, furniture). This is not a necessity but enables us not to consider the loop-closing problem in mapping, and gives a predefined configuration of the environment of the robot (a static map of the residential place of the robot).

The field of perception for each sensor is different, as shown in figure 2. The field of vision, represented by the purple cone in figure 2(a) is narrower than the field of audition, represented by the purple zone in figure 2(b), which means that a person can be heard without being part of the field of vision.
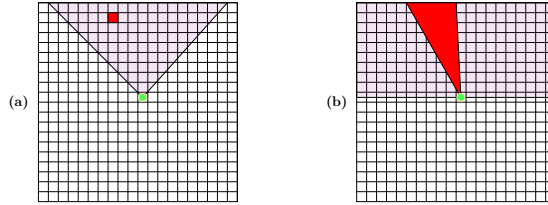

Fig. 2: Perception characteristics: In both maps the robot is represented by a green point and is assumed to be facing north. In (a), the field of vision is represented by the purple zone. When a human or an object detection happens (in red), it can directly be put at its precise location inside the grid. In (b), the field of audition is represented by the purple zone. A detection (in red) is less precise. The source can be anywhere in the red zone.

We assume that human detection is performed on each frame of the video (e.g. face or skeleton detection). On each frame, if a human is detected, its position in the image is known. A video observation cannot directly be placed in a grid, as the depth information is missing. In order to be able to localize video observation inside sensor grids, a registration is performed between the information provided by the RGB sensor and the point cloud given out by the depth sensor. This registration enables the linking between each pixel of the video and corresponding cells inside the grid (i.e. correspondance between the image frame of reference and the world frame of reference). Thus thanks to the registration, any video observation can be placed inside the sensor grids, as shown in red in figure 2(a). In the same way, we assume that a human sound-source localization is performed on the audio in synchronization with the video (typicall thanks to a generalized cross-correlation method), so that each audio observation is made at the same time as the video observation. As the robot

only has two microphones, the localization cannot yield a precise position of the source but a cone in which the source is likely to be, as shown in red in figure 2(b).

3.2   Sensor grids

An evidential grid is created for each sensor. The aim of evidential sensor grids is to fill each cell of the perceived spatial zone with a belief function on discernment space $\Omega = \{H, O, F\}$, described in section 2, matching the information extracted from the sensor for this particular cell.

Video evidential grid: As it was said above, each video observation can be placed inside the grid through registration. The video itself can only deliver information on the possible presence of a human in the scene, while the depth sensor, through the point cloud gives information on the presence of objects (human or not) and free space. This leaves four possibilities for a given cell:

- If a human is detected a bba is created in the matching cell with the following focal elements : H [1] and Ω
- If an object is detected by the depth sensor without corresponding to a human detection, a bba with focal elements $\{H, O\}$ and Ω is created. Indeed, the depth sensor does not give information as to what the nature of the detected object is.
- If neither the depth sensor nor the video detect an object, a bba with focal elements F and Ω is created.
- If a spatial zone is not part of the field of vision of the robot or is occluded because of an object, a vacuous belief function is created for each cell of this zone.

The way the belief are attributed reflects the trust in the detection algorithm.

Audio evidential grid: As shown in figure 2(b), sound-source localization with two microphones is not spatially precise. The sound source can be anywhere in the cone. Moreover, the sound can only give out information on the possible presence of a human in a given cell (not on the fact that a given cell is free space). Thus for the audio perception map, two cases are possible for a given cell:

- If a detection happens and the considered cell is part of the audio cone, a bba with focal elements H and Ω is created.
- If the considered cell is not part of the audio cone or if no detection is made (silence), a vacuous bba is created for the cell.

In case of a detection, all the cells in the spatial zone of detection receive the same bba. The way the beliefs are attributed reflects the trust in the detection algorithm.

[1] In order to simplify the notations, singletons of the discernment frame are noted without brackets, hence H instead of {H}

4   Fusion scheme

4.1   Perception grid

At each observation time, both an audio and a video evidential grids are created. In order to obtain a unified perception map accross all sensor, the audio and video perception map need to be fused. The grids are fused cell by cell, thanks to an unnormalized conjunctive combination. However, the detections do not have the same spatial uncertainty, depending on the modality, which means that conflict can appear during fusion. Knowing the possible focal elements of the belief functions, only one combination can create conflict, as shown in table 1: it is the case where a cell is part of the audio cone, which mean that there is a possible presence of a human in this cell, whereas the vision sensors detect the cell as being free.

m_a^{Ω} \ m_v^{Ω}	H	$\{H,O\}$	F	Ω
H	H	H	$\emptyset$	H
Ω	H	$\{H,O\}$	F	Ω

Table 1: Intersections between possible focal elements of the audio belief function m_a^{Ω} and the video belief function m_v^{Ω}. The states are : F for free space, H for human presence, and O for non-human object

In that case we chose to assign the conflict to hypothesis F, which means trusting the vision over the audio detection, as the vision is spatially more precise. The fused basic belief assignment m_P^{Ω} is then obtained through this modified fusion rule:

$$m_P^{\Omega}(S) = \left(m_a^{\Omega} \textcircled{0} m_v^{\Omega}\right)(S), \quad \forall S \subseteq \Omega, S \neq F \tag{1}$$

$$m_P^{\Omega}(F) = \left(m_a^{\Omega} \textcircled{0} m_v^{\Omega}\right)(F) + (m_a^{\Omega} \textcircled{0} m_v^{\Omega})(\emptyset) \tag{2}$$

After this fusion, we obtain the perception grid, representing the information from all sensors, in which each cell contains a basic belief assignement m_P^{Ω}, with four possible focal elements: $H, \{H,O\}, F, \{\Omega\}$.

4.2   Temporal fusion

We would like the framework to take the dynamicity of the scene into account. Those kind of information are given through temporal integration. A type of information that can be useful to a companion robot is to differentiate between mobile and static objects, e.g. a robot will not behave in the same way with a mobile human and with a sitting human.

In order to be able to differentiate between static and mobile objects, we introduce the frame of discernment $\Theta = \{H_s, H_m, O_s, O_m, F\}$ where the new states are:

- H_s: Presence of a static human, i.e. a human has been occupying this cell for an arbitrary period of time
- H_m: Presence of a mobile human, i.e. a human has filled the cell recently
- O_s: Presence of a static non-human object, i.e. an object has been occupying this cell for an arbitrary period of time
- O_m: Presence of a mobile non-human object, i.e. an object has filled the cell recently
- F: free space

To be able to work in this discernment space we define the following refining:

$$\gamma: \begin{array}{ccc} \Omega & \to & 2^{\Theta} \\ \{H\} & \mapsto & \{H_s, H_m\} \\ \{O\} & \mapsto & \{O_s, O_m\} \\ \{F\} & \mapsto & \{F\} \end{array} \tag{3}$$

$$\begin{array}{ccc} 2^{\Omega} & \to & 2^{\Theta} \\ m_P^{\Omega}(B) & \mapsto & m_P^{\Theta}(\gamma(B)), \quad \forall B \subseteq \Omega \end{array} \tag{4}$$

This refining is applied to each cell of the perception grid, which enables us to obtain a grid with discernment frame Θ. The sensor grid cannot add belief on dynamic hypotheses as sensors are not able to detect motion in the scene. This can be done by detecting changes in the scene between the time t-1 and time t. In order to do that, we propose the use of a credibilist hidden markov model.

Credibilist Hidden Markov Model: A credibilist HMM estimates a belief function at time t from the estimate of this belief function at t-1 and observations at time t.
To be able to perform the HMM, two elements are needed:

- $q_t^{\Theta}[S_i^{t-1}](S_j^t) \ \forall S_j, S_i \subseteq \Theta$, is the conditional commonality of transition from credibilist state S_i at t-1 to the state S_j at time t. Thus the matrix q_t^{Θ} is a matrix of size $2^{|\Theta|} * 2^{|\Theta|}$
- $q_b^{\Theta_t}(P) \ \forall P \subseteq \Theta$ which represents the commonality repartition on subsets of Θ given by observations at time t

With those elements, the estimation is a two-step process. First the commonality function at time t $\widehat{q}_{\alpha}^{\Theta_t}$ is predicted:

$$\widehat{q}_{\alpha}^{\Theta_t}(S_j^t) = \sum_{S_i^{t-1} \subseteq \Theta_{t-1}} m_{\alpha}^{\Theta_{t-1}}(S_i^{t-1}) \cdot q_t^{\Theta}[S_i^{t-1}](S_j^t), \quad \forall S_j^t \subseteq \Theta_t \tag{5}$$

where $m_{\alpha}^{\Theta_{t-1}}$ is the estimated basic belief assignment at time t-1.

The estimated commonality is then corrected with the observations at time t, by conjunctive combination:

$$q_{\alpha}^{\Theta_t}(S_j^t) = \widehat{q}_{\alpha}^{\Theta_t}(S_j^t) \cdot q_b^{\Theta}(S_j^t), \quad \forall S_j^t \subseteq \Theta_t \tag{6}$$

Prediction step: In our case, no model of evolution of a cell is available, as predicting the way each cell of the scene is going to change is complicated. That is why we propose to perform a temporal discounting on each cell of the estimated grid at t-1 as a prediction. This enables the model to slowly decrease towards ignorance in case no further observation is made on the cell :

$$\widehat{m}_\alpha^{\Theta_t}(S_i) = m_\alpha^{\Theta_{t-1}}(S_i) \cdot e^{-\gamma} \quad \forall S_i \subseteq \Theta_{t-1} \tag{7}$$

$$\widehat{m}_\alpha^{\Theta_t}(\Theta) = 1 - \sum_{S_i \subseteq \Theta_{t-1}} \widehat{m}_\alpha^{\Theta_t}(S_i) \tag{8}$$

where γ is the decrease rate toward ignorance. This is equivalent to defining:

$$m_t^\Theta[S_i^{t-1}](S_j^t) = 0, \quad \forall i \neq j, S_j^t \neq \Theta \tag{9}$$

$$m_t^\Theta[S_i^{t-1}](S_i^t) = e^{-\gamma}, \quad S_i \neq \Theta \tag{10}$$

$$m_t^\Theta[S_i^{t-1}](\Theta) = 1 - e^{-\gamma}, \quad \forall S_i^{t-1} \neq \Theta \tag{11}$$

$$m_t^\Theta[\Theta](\Theta) = 1 \tag{12}$$

Correction step: For each cell, the correction step is supposed to be a conjunctive combination between $\widehat{m}_\alpha^{\Theta_t}$ and $m_P^{\Theta_t}$. However, each time the fusion is performed, some conflict will appear if the cell's true state changes between t-1 and t, as the model does not predict any change. Information must be extracted from the conflict. Once again the only case where conflict appears is when a cell change state from occupied (by a human or an object) and free, or conversely when it becomes occupied. In the former case, the only viable option is to trust the sensors and transfer the conflict on the hypothesis F (trusting the prediction step would be dangerous as it does not account for any possible state change in the cell). However, in the latter case, the cell filling with an object means that a mobile object just entered the cell, and we propose to transfer the conflict on the underlying mobile hypothesis of $m_P^{\Theta_t}$. This corresponds to the following rule of fusion:

$$m_\alpha^{\Theta_t}(A) = \left(\widehat{m}_\alpha^{\Theta_t} \circledcirc m_P^{\Theta_t}\right)(A) + \sum_{\substack{A=(B \cap \{H_m, O_m\}) \\ B \subseteq \Theta \setminus F}} \left(\widehat{m}_\alpha^{\Theta_t}(F) \cdot m_P^{\Theta_t}(B)\right), \quad \forall A \subseteq \{H_m, O_m\} \tag{13}$$

$$m_\alpha^{\Theta_t}(F) = \left(\widehat{m}_\alpha^{\Theta_t} \circledcirc m_P^{\Theta_t}\right)(F) + \sum_{A \subseteq \Theta \setminus F} \left(\widehat{m}_\alpha^{\Theta_t} \cdot m_P^{\Theta_t}(F)\right) \tag{14}$$

$$m_\alpha^{\Theta_t}(A) = \left(\widehat{m}_\alpha^{\Theta_t} \circledcirc m_P^{\Theta_t}\right)(A), \quad \forall A \ else \tag{15}$$

Staticity check: At that point, the system never adds belief on static hypotheses H_s and $\{H_s, O_s\}$. It can also artificially increase the belief on mobile hypotheses in the case where a previously free cell is filled, e.g. by a human, and

the object that fills it stops: in this case, as the sensor continue to detect the human (focal element $\{H_s, H_m\}$), when the temporal fusion happens, the mass will be transferred on H_m for as long as the person stays in the cell, because of the conjunctive combination. To answer this problem, we propose a staticity check, which is done after each temporal fusion. This test enables the transfer of masses on static hypotheses in case a cell has been occupied for a predetermined time. A cell is considered occupied if the belief in the hypothesis $\Theta \setminus F$ is larger than a predefined number, i.e. if the percentage of belief in occupation is higher than an arbitrary threshold.

The occupation is defined as follow :

$$Occ = bel(\{H_m, H_s, O_m, O_s\}) \tag{16}$$

if Occ is larger than a predefined δ for a time longer than Δt then all masses on mobile hypotheses are transferred on static hypotheses i.e. :

$$m(\{H_m\}) \mapsto m(\{H_s\}) \tag{17}$$

$$m(\{H_m, O_m\}) \mapsto m(\{H_s, O_s\}) \tag{18}$$

$$m(\{H_m, O_m, H_s, O_s\}) \mapsto m(\{H_s, O_s\}) \tag{19}$$

By doing this, if a mobile object suddenly stops, and stays detected, the belief will be naturally transferred on static hypotheses.

5   Results

This section shows the performances of the whole framework presented in the previous sections. The algorithm is shown in a simulation framework, in order to be able to fully control the parameters.

5.1   Simulation Conditions

In the simulation, an occupancy grid is available, showing the true state of the scene, of size 50x50 cells. With cells of size 20x20cm, this represents a place of $100m^2$. In order to properly illustrate the fusion scheme without overcomplicating the analysis, we propose a simplified environment for the robot, shown in figure 3. The environment is composed of free space and an obstacle, e.g. a wall.

The robot occupies one cell at a time. This is not restrictive as the robot is not part of its own perception. It can, however, be of importance for the robot to know its own volume when it moves in the scene, in order to avoid collisions. We assume that humans of the scene also occupy one cell at a time. This might cause some problems locally as humans are never truly motionless, but as it depends of the spatial resolution of the grids, we consider it a sufficient first approximation.

As for the perceptual capabilities of the robot, they are shown in figure 3b and 3c. We used a cone of visibility of angular size 45 degrees centered on the robot, and an audio cone of angular size 20 degrees centered on the observation to account for audio localization uncertainty. We assume that the robot is able to localize itself properly in the map, which is not unrealistic, considering the performances of available state-of-the-art localization algorithms [9].

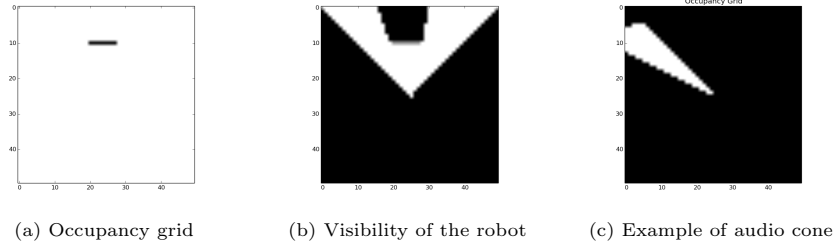(a) Occupancy grid   (b) Visibility of the robot   (c) Example of audio cone

Fig. 3: Occupancy grid and perceptual capabilities of the robot. 3a shows the occupancy grid: white cells represent free space and black cells represent obstacles. 3b shows the visual perception space of the robot. In this case the robot is placed in the center of the grid, facing north. White cells represent places the robot can see, black cells cells that are not visible. 3c shows an audio detection by the robot (white cells).

5.2 Scenario

In the scenario described in this section, the robot does not move. This is not restrictive as we assume that the robot localizes itself properly, as the state of each cell is estimated independently of the state of its neighbourhood. We consider a scenario containing one moving and talking person, e.g. on the phone. The person crosses the field of vision of the robot. His voice may continuously be heard. The steps are the following ones:

1. When the robot starts perceiving, the human is outside its field of vision, but inside its audio cone detection, talking
2. The human starts to move and enters the field of vision of the robot
3. The human continues walking in front of the robot and becomes silent
4. The human stops for a while and stand without moving, e.g. listening

   The parameters used are:

   – $\gamma = 0.4$: we use a rather slow decrease rate, in order to see the influence of temporal discounting
   – $\Delta t = 6$: after 6 consecutive observations of occupancy, the staticity check will perform on the cell. This leaves a good amount of time before adding belief to staticity
   – $\delta = 0.5$: if there is more belief that support the occupied state of the cell than the free state, we consider the cell to be occupied

   In order to present the information in a compact way, only the subset of Θ with the highest assigned belief is shown for each cell. The three first step of the scenario can be observed on figure 4.

   At t=0 (figure 4a and 4d), the robot begins to perceive its environment. As there is no previous fused map, the sensor grid is the final map. The human is not inside the field of vision, and the audio cone is filled with belief functions on $\{H_s, H_m\}$. As expected, the obstacle in the field of vision receives belief functions with belief on $\{H_s, H_m, O_s, O_m\}$. The rest of the space is accurately separated between free space in the rest of the field of vision and ignorance on

(a) Occupancy grid at t=0   (b) Occupancy grid at t=5   (c) Occupancy grid at t=9

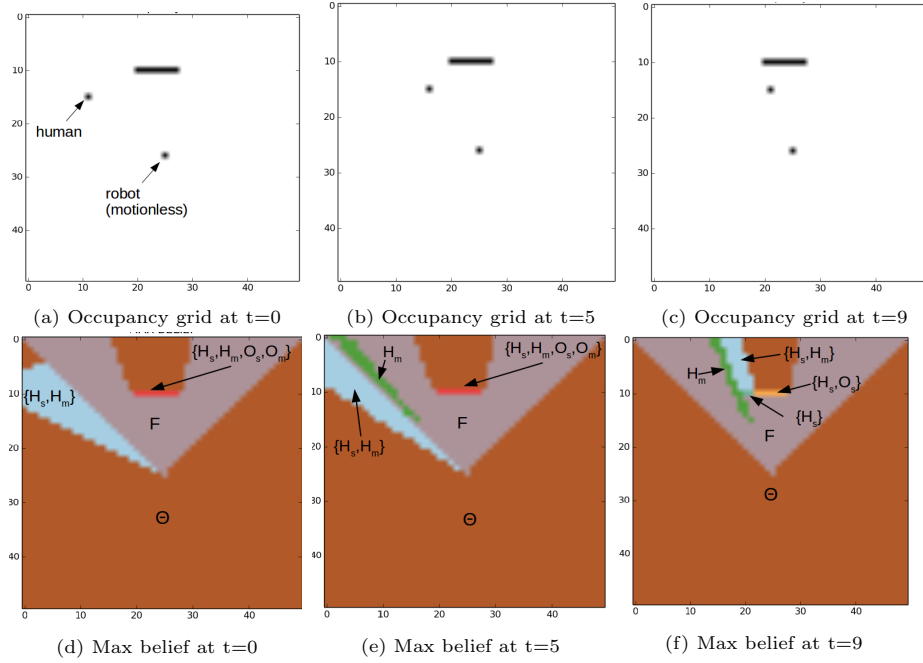(d) Max belief at t=0   (e) Max belief at t=5   (f) Max belief at t=9

Fig. 4: Scenario realization: on the top line, occupancy grids at 3 different observation time, and on the bottom line, corresponding max belief grid after fusion. On each zone, the subset of Θ with the maximum belief assigned is indicated.

non-perceived zones.

At t=5 (figure 4b and 4e), the human has entered the cone of vision, and is still talking. The temporal discounting leaves a part of previous audio cones on the left (hypothesis $\{H_s, H_m\}$). As the human is walking, both the audio cone and the video detection are changing at each observation, filling cells that were previously considered free, thus the presence of belief functions with maximal belief on H_m. As expected, the case of conflict between video putting belief on F and audio putting belief on $\{H_s, H_m\}$ is solved by trusting the video, hence the fact that the audio cone is not appearing once the human enter the field of vision.

At t=9 (figure 4c and 4f), the human continues walking and talking and is now walking in front of the obstacle. As before, in the visually perceived zone, the hypothesis H_m is dominant in cells that are becoming occupied from previously free. This time the audio cone has two different kinds of impact: on the zone behind the obstacle, it is fused with ignorance, hence the dominance of hypothesis $\{H_s, H_m\}$, which seems fair, as the sound might in fact come from behind the obstacle with a non-visible person. Interestingly, the cone also has an influence on the obstacle itself, as the part of the obstacle that is inside the audio cone will receive a belief function on $\{H_s, H_m\}$ (due to the fact the audio is increasing the belief in the presence of a human and the video cannot contredict the informa-

tion). However as those cells have been occupied for longer than 6 consecutive observations, the belief is transferred on $\{H_s\}$. This can be an advantage, as there are a lot of common obstacles on which a human can be (tables, chair, sofas, etc..). However, to obtain a more accurate state estimate, a good idea would be to include information from the occupancy grid (e.g. a cell containing a wall cannot receive belief on H). The staticity check also happens on the rest of the obstacle, which transfers the belief from $\{H_s, O_s, H_m, O_m\}$ to $\{H_s, O_s\}$.

(a) Occupancy grid at t=14   (b) Occupancy grid at t=25
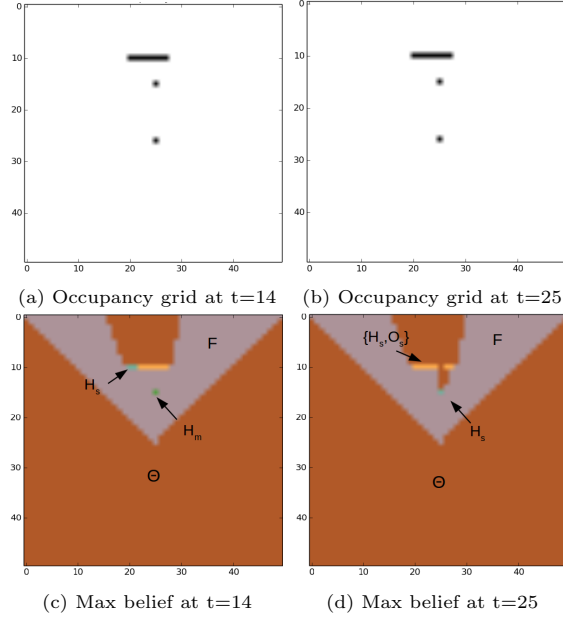
(c) Max belief at t=14   (d) Max belief at t=25

Fig. 5: End of the scenario

The rest of the scenario can be seen on figure 5.
At t=14, the person has stopped talking and stops moving. Not enough time has passed for the staticity check to enter into effect. Moreover the temporal discounting, linked with the video observation of the obstacle has not allowed yet to go back to the dominance of $\{H, O\}$ on the totality of the obstacle. Finally, at t=25, the staticity has had effect on the human detection and the cell containing the person is accurately dominated by a belief on H_s. The obstacle is detected as such again, and the temporal discounting created a zone of ignorance behind the human. The filter describes the environment of the robot accurately.

6   Conclusion and perspectives

This paper presented an evidential filter for indoor navigation of a mobile robot in dynamic environment. The filter is based on two main steps: the creation of evidential grids to represent sensor data, and the fusion scheme divided into two

parts. First the sensor data is fused accross all sensors and then the information is integrated inside an evolutive model. The filter describes the environment of the robot accurately extract accurate information from conflict between the sensors and in the evolutive model.

One interesting addition could be the use of neighbourhood to estimate the sate of a cell: this would enable tracking and a more elaborate model (spatially coherent). Moreover it would help face the possible loss of detection from sensors.

References

1. Chrungoo, A., et al.: Towards perception driven robot-assisted task-oriented therapy. In: ICORR (2015)
2. Elfes, A.: Using occupancy grids for mobile robot perception and navigation. Computer (1989)
3. Kurdej, M., et al.: Map-aided fusion using evidential grids for mobile perception in urban environment. In: Belief Functions: Theory and Applications. Advances in Intelligent and Soft Computing (2012)
4. Labourey, Q., et al.: Audiovisual data fusion for successive speakers tracking. In: VISIGRAPP (2014)
5. Moras, J., et al.: Credibilist Occupancy Grids for Vehicle Perception in Dynamic Environments. In: ICRA (2011)
6. Ramasso, E.: State sequence recognition based on the Transferable Belief Model. Theses, Université Joseph-Fourier - Grenoble I (2007)
7. Siagian, C., et al.: Autonomous mobile robot localization and navigation using a hierarchical map representation primarily guided by vision. J. Field Robotics (2014)
8. Smets, P.: Belief functions: The disjunctive rule of combination and the generalized bayesian theorem. In: Classic Works of the Dempster-Shafer Theory of Belief Functions (2008)
9. Thrun, S., et al.: Probabilistic Robotics. The MIT Press (2005)
10. Zaraki, A., et al.: Designing and evaluating a social gaze-control system for a humanoid robot. Trans. on Human-Machine Systems (2014)