A lidar Perception Scheme for Intelligent Vehicle Navigation

Julien Moras, Véronique Cherfaoui, Philippe Bonnifait

A lidar Perception Scheme for Intelligent Vehicle Navigation

Julien Moras, Véronique Cherfaoui and Phillipe Bonnifait
UMR CNRS 6599 Heudiasyc
Université de Technologie de Compiègne
Compiègne, France
first-name.name@hds.utc.fr

***Abstract*—In urban environments, detection of moving obstacles and free space determination are key issues for driving assistance systems or autonomous vehicles. This paper presents a lidar-based perception system for passenger-cars, able to do simultaneously mapping and moving obstacles detection. Nowadays, many lidars provide multi-layer and multi-echo measurements. A smart way to handle this multi-modality is to use grids projected on the road surface in both global and local frames. The global one generates the mapping and the local is used to deal with moving objects. An approach based on both positive and negative accumulation has been developed to address the remnant problem of quickly moving obstacles. This method is also well suited for multi-layer and multi-echo sensors. Experimental results carried out with an IBEO Alasca and an Applanix positioning system show the performance of such a perception strategy.**

***Index Terms*—Intelligent Vehicles, Environmental Perception, Mapping, Mobile Object Detection**

I. Introduction

Autonomous vehicles are becoming a reality in urban areas for human transportation. Indeed, several works in the world have shown some impressive results. However, autonomous driving in urban environment remains a problem and need the understanding of the scene to predict its evolution. Perception systems use different sensors and the measurements can be noisy, biased or incomplete. New lidar technologies (multi-layer, multi-echo) can bring a solution to this problem by increasing significantly the number of measurements. This sensing technology provides reliable perception of the surroundings even if the laser beam is partially reflected. The perceptive problem can be decomposed in different parts: the localization, the mapping and the mobile objects detection and tracking.

- Ego-Localization (EL) :
  Positioning task is a recurrent problem in robotic applications. GPS, Inertial Measurement Units (IMU), proprioceptive or exteroceptive odometry are main solutions proposed for Inteligent Vehicles. Today, reliable submetric positioning systems exist but they remain too expensive for commercial application.
- Mapping (M):
  Many robotic works have treated the mapping of static environments with different approaches. Generally, mapping is coupled with the localization task to address the Simultaneous Localization And Mapping (SLAM) problem [1], [2], [3]. Two main approaches exist: the feature-based approach tries to map with a predetermined set of geometrical shapes (segment, arc, etc...) and the grid approach is based on a discret space representation that makes easier data association and fusion.
- Mobile Object Detection and Tracking (MOT) :
  SLAM methods are very sensitive to the presence of moving objects in the scene. Indeed, the algorithms are based on the temporal coherency of the mapping process. A way to address this issue is to detect and track the moving objects [4], [5], [6], [7]. Usually, a detection and tracking system is developed in the feature-based framework and works in 3 steps: clustering of raw data for object detection, data association and temporal fusion. The main sources of error are in the clustering phase and in the association step. Some recent works use parametric models [8]. Using a grid approach for the MOT problem is not usual but some works relative to the Bayesian Occupancy Filter (BOF) [9] tend to solve that by clustering cells with specific criteria.

In this work, we consider a mobile robot moving in planar world. As we assume to be in an urban environment, there can be a lot of obstacles and a lot of moving objects as illustrated in Figure 1. In this paper, we do not consider the ego-localization (EL) problem, that's why we use a positioning system 'Applanix Pos LV 220' which is an integrated multi-sensor system that provides a very precise 3D pose (position and heading) of the vehicle. The two other problems (M and MOT) are linked together because they deal with the perception system. The main difference is the mobility of the objects. It is the criterion that we use to do the classification. To manage this strategy, we use the grid-based framework.

This paper presents a perception scheme mainly dedicated to the detection of the mobile objects on the souroundings allowing the mapping of the static environement in spite of the presence of moving objects. This detection strategy can be used afterward in a tracking algorithm to be able to predict the free space for navigation applications. The main advantage of this approach is that there is no clustering phase and it is adapted to a wide range of urban obstacles (pedestrian, vehicle, bicycle,...). It can also cope with the multi-echo measurements that are frequent with new generation lidars.

In the first section, the framework to fuse sensors data is presented. Then, the combination strategy based on accu-

978-1-4244-7815-6/10/$26.00 ©2010 IEEE

mulation will be described followed by a section showing experimental results.

Fig. 1. Urban situation: the red car is equipped with a multi-layer lidar

II. Fusion framework

We use a 2D grid representation that is defined as a discret space projected on the floor. Each cell of the grid represents a piece of space and contains data computed from perception of the environment. We use a dual space representation: an instantaneous local map called Scan-Grid (SG) which is composed of polar cells of size $L \times \varphi$ (L is a length and φ represents an angular sector) and a global map called Grid-Map (GM) which is Cartesian map referenced in the world. Cartesian cells are L-length squared. This section describes how these maps are defined according to the sensor model and the environment model.

A. Lidar Scan

The sensor used is an IBEO Alasca lidar. It is a four-scan sensor which provides a 3D points cloud of the environment at each scan. This sensor can do measurements up to 200 meters in a front field of 320°, with a rate from 8Hz up to 40hz depending on the needed angular resolution. It uses a 905 nm wavelength Infra-Red laser which has an aperture of 0.25°. The angular resolution is adaptive according to the angle as shown in figure (2). This sensor is also able to provide several echoes per line of sight if, for instance, the laser beam is partially reflected by an obstacle. Another characteristic is that it can return no measurement in the considered line of sight. If there is no echo, two cases are possible : there is no object until the max range or there is is an obstacle which doesn't reflect the laser beam. Therefore, the sensor model presented hereafter, proposes to take into account of these sensor particularities.

B. Multi-echo Scan-Grid

Using the sensor model, we can build from each lidar scan, the Scan-Grid (SG), which is a 2D local instantaneous occupancy grid . Because the lidar is a polar sensor, a polar grid model is used to compute the occupancy of the cells. The sensor precision is higher than grid resolution and the grid is sensor-centered. So, we can use an ideal sensor model as a 1D explicit solution as shown on the top part of the Figure 4. Considering the Scan-Grid, let $SG(r, \theta)$ denote the

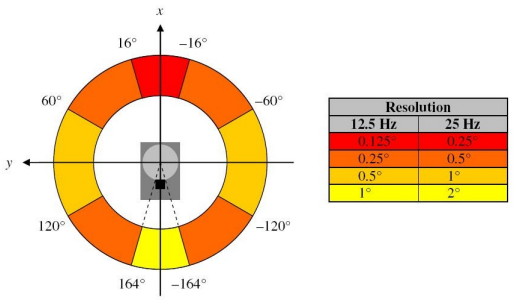

Fig. 2. Alasca XT angular resolution in function of the angle of measurement and of the frequency

state of a cell taking integer value, refering to three different states for the cell described by the table I. The intensity of the value corresponds to the redondancy, that is a mesure of the confidence.

Cell State	SG values
occupied	positive
unknown	null
free	negative

TABLE I
SCAN-GRID VALUES TABLE

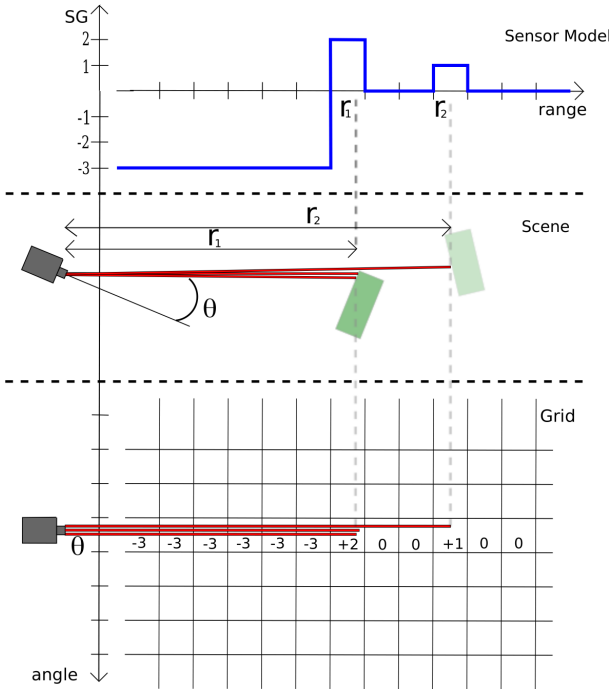

Fig. 3. Multi-echo sensor model. Example of one row of the Scan-Grid computation : in the middle, a bird view of the real scene. On the top, the figure shows the sonsor model compute from the multiple echos as described in .On the bottom, figure shows the affectation for the row corresponding to this angle.

This is illustrated by the Figure 3. Each row of the scan grid corresponds to one angular sector. In this sector, several

echoes are possible because of three reasons: i) in one direction it is possible to receive several echoes, ii) the projection of 4 layers on the same plane can provide echoes located at different distances and iii) the angular resolution of the lidar is not constant and some times better than the resolution of the grid (several lidar directions can be projected on the same column of the polar grid). Therefore, it is worth to notice that this grid sensor model takes into account the multiple echo capabilities. The SG is initialized with 0, each angle is processed independently as shown by the Figure 3, each measurement increases the value of the corresponding cell and decreases the value of the cells preceding the first occupied cell. The negative information characterizes the free space. Cells between two occupied cells and cells following the last detected cell are affected at 0. An example of SG is given in Figure 4 where multi-echoes are observed.

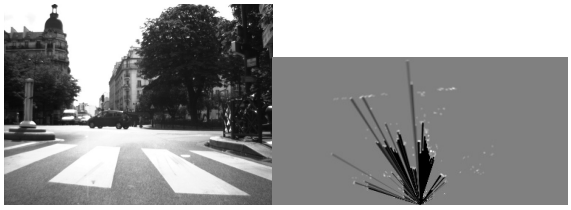

Fig. 4. Multi-echo sensor model: on the left the camera view of the scene and on the right the SG projected in a Cartesian frame, white cells are occupied, black are free, and gray are unknown.

C. The Grid-Map

The Grid-Map (GM) is defined as a global accumulation 2D grid and stores the mapping information. It is a cartesian map and each cell of size $L \times L$ is a piece of the 2D projection of the global space. By global, we mean that this grid is referenced to an East-North-Up frame which is considered fixed. The GM is used to create the lidar based perceptive map of the area. This implies that all the static elements of the scene will be mapped included pseudo static elements such as parked cars.

D. From the Scan Grid to the Grid Map

A key point of the method is that the SG can be projected in the GM frame using the pose provided by the Applanix sensor, this is illustrated by Figure 5. First, the polar grid is converted in a Cartesian one using a bi-linear interpolation. Then, a transformation of the Cartesian SG is applied in order to make the projection. This consists in one rotation and one translation. The rotation is done with a bi-linear transformation, because each projected cell may project partially on several cells. Bi-linear transformation may interpolate values, so, in the transformated cell, value are mixtures between the values of the neighborhood of the polar cell. This causes a smoothing of edges.

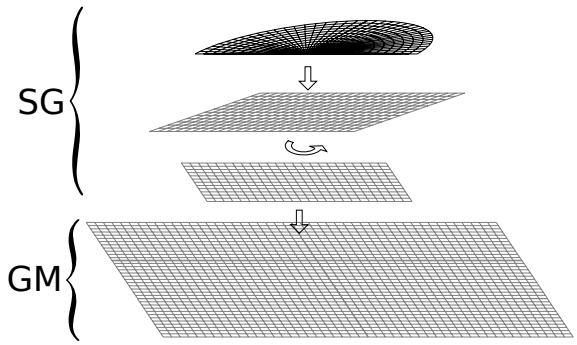

Fig. 5. Successive transformations to use the SG with the GM

III. Implementation of the Accumulation Algorithm

The processing architecture applied to the GM is described in Figure 6. This section details the GM updating phase with an accumulation algorithm. The framework presented in the section 2 is generic and instead of making the fusion using the presented accumulation algorithm, a probabilistic or evidential fusion framework could be used. The accumulation process is described to address the problem of Mapping and Mobile Objects Tracking (MMOT).

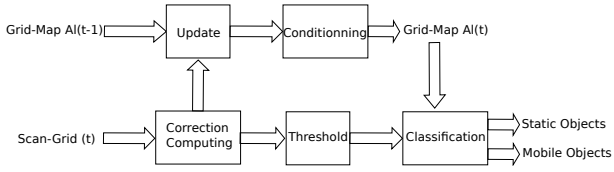

Fig. 6. Architecture of the GM process

A. Negative and positive accumulation in the GM

In an accumulation strategy, each cell of the GM stores an indicator Al which represents the current accumulation level of occupancy.

$$Al \in [A_{min}\, A_{max}]$$

The limit A_{min} represents a free cell whereas A_{max} represents an occupied cell.

This map is initialized with the average value $A_0 = \frac{A_{max}+A_{min}}{2}$ in all the cells. This means that there is no a prior knowledge of the environment.

The update process provides a positive or negative accumulation using the values coming from the SG. The incrementing process aims to build the map by integrating occupancy measurements like a 2D histogram [10]. Some recent works have used a similar approach applied to the SLAM problem [11]. This work focus on the localization problem, using matching on the map to realize odometry. In opposition, in our method, localisation is provided, there is no matching or data association since the algorithm just performs a filtering in order to extract moving objects.

The mapping step is composed of 3 stages.

- Correction computing

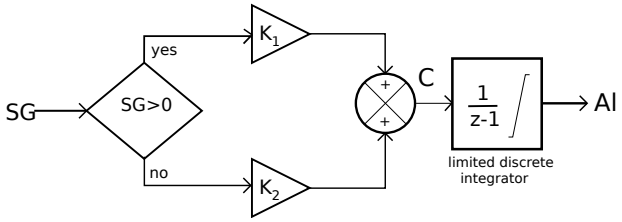

Fig. 7. Positive and negative accumulation mechanism in a cell of the map

Using the SG projected in the GM frame, the correction term c_t is calculated according to the sign of the cell value (Fig. 7).

First, the accumulation levels of the occupied cells are incremented using the K_1 gain. This carry out the mapping of the static world, but the cells crossed by mobile obstacles are incremented too. The principal difference between a static and a moving obstacle is that the moving one does not occupy the same position all the time. The occupied cell will become free after the moving object leaves the cell.

That's why the empty cells are decremented using the K_2 gain. This negative accumulation process has a non negligible response time. In order to reduce this lag, K_2 has to be larger than K_1. A particular problem of this solution is that small obstacles which are not detected at each scan will be not mapped.

- Updating

The updating process is an integration. It is realized by summing the accumulation level with the correction term:

$$Al_t = Al_{t-1} + c_t$$

Where Al_t and c_t are respectively the Accumulation level and the Correction at the time t as show on Figure 7.

- Conditioning

In this integration method, the level of accumulation depends of several parameters : occupancy, integration time and visible ratio of the object with respect to the size of the cell. To solve this, a saturation limits the accumulation level. Figure 8 shows the accumulation level behavior of one cell according to the SG value.

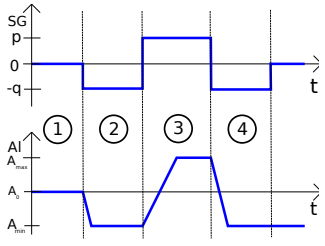

Fig. 8. In part 1, the cell is not scanned: Al doesn't change. In part 2, the cell is scanned as empty: Al decreases until threshold A_{min}. In part 3, the cell is detected as occupied: Al increases with a different speed to A_{max}. In part 4, the cell is declared empty after few iterations.

B. Mobile Object Detection

The detection method we propose consists in comparing the current SG with the GM. Because of the transformations (polar/cartesian and local/global), the projection of the SG on the GM does not contain integer values but real ones. SG cells are compared with a detection threshold D_{Th} in order to consider only the cells containing an object. Then, the Al of corresponding GM cell is compared with a classification threshold C_{Th} in order to define if cells are occupied by moving or static objects. Figure 9 explains the classification scheme.

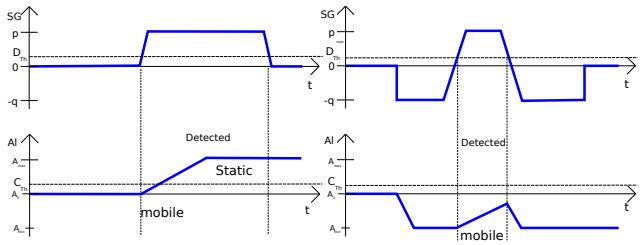

Fig. 9. Behavior of the Al of an occupancy cell depending on the dynamic of the object. On the left, a static object reaches the threshold after a few time.On the rigth, the cell occupied by a moving object does not reach the threshold C_{Th} during the time the cell was detected.

IV. Results

A. Experimental setup

The algorithm presented before was tested on a data-set acquired with a real vehicle in urban conditions. The complete data set is a 20-minutes long sequence, acquired in collaboration with the French geographic institute (IGN) in Paris. The vehicle follows a reference track (Fig.10) repeated 3 times.

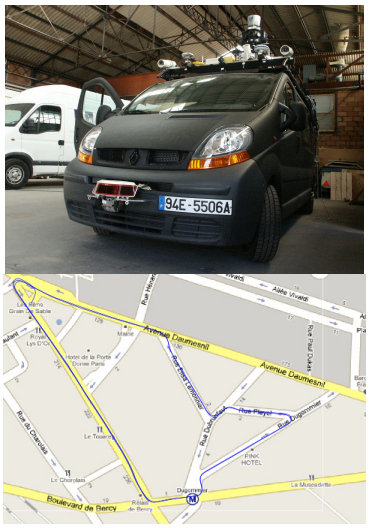

Fig. 10. On the top the test vehicle with the lidar sensor in front. On the bottom, the track follow by the vehicle

We exploit the data provided by 3 sensors: the lidar Alasca XT, the Applanix positioning system and one camera. In order to synchronize the data between the sensors, time stamps in GPS time have been used. The lidar was mounted in a way to have its lowest layer horizontal. Its acquisition frequency

was 15 Hz, with an angular resolution max of 0.25° in front of the vehicle and 1° on the side. The Applanix positioning system is composed by 2 GPS, 1 IMU and 1 odometric sensor. The quality of the precision has been evaluated by comparing its output with 2 others positioning systems IXSEA LandINS (high precision positioning solutions than are usually used as ground truth). The camera was triggered by the lidar: this simplifies the synchronization problem. This camera was installed just under the lidar at the front bumper. Until now, the camera is just used for validation and scene visualization.

The SG covers a range of 200 meters and angulare field of 180°, with a resolution of 0.5m in range and 1° in angle. The global GM used covers 800m x 700m area with a resolution of 0.5m.

The algorithm has been implemented with MATLAB. It is not real time in the current implementation: it takes 1.5 s on a laptop (Intel Centrino at 2Ghz) to compute and display one step.

In this implementation, the gain K_1 is set to 1 and the Al takes value in $[0\,30]$, that implies a maximal lag of 2s for static elements mapping in the case they are detected at each scan. In order to increase the reactivity of the system for mobile object mapping, the decrementing gain K_2 has to be higher than K_1. K_2 is set to 5 in this experiments. We haven't choose a too important value to avoid that wrong measurement affects significantly the map. The detection threshold D_{Th} was set to 0.5 and the classification threshold C_{Th} was set to 10. All these values were fixed in an trial-and-test way.

B. Results

The validation was made in two steps. First, the validation of the mapping and then the validation of the classification.

The mapping validation has been done using a referenced free space map provided by the IGN. This map does not take into account of the presence of some long time stationary objects like cars parked.

The classification validation is made by using images coming from the camera. It is illustrated here using several scenarios (intersection, crossing car,etc..) detailed afterward.

1) Mapping Results: We compare the map built during one loop of the reference track (Fig. 11) with the the map called "free space map" provided by the IGN. This free space map is a 3D mesh of the road surface, it was created manually by expert operators from high-resolution aerial images using photogrametry.

The key point we want to put in is that in spite of many moving objects were present in the scene, static object like buildings are correctly mapped and the free space is correctly determined. Depending of the sensor field of view, we can map several aligned objects like parked cars and building using different lidar layers. Since moving objects are not tracked in time and can momentarily become static (by stopping at traffic light for example), they can appear as static environement in the GM. This problem appears in the partircular case where mobile objects leave the field of view of the ego-vehicle sensor when they are stoped. Small objects (typically less than 0.5m

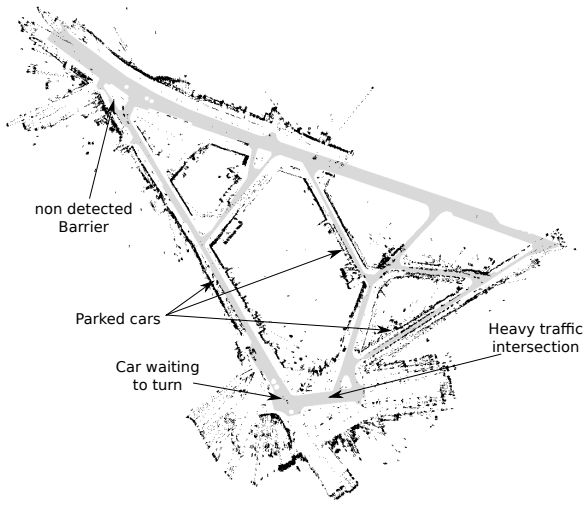

Fig. 11. Comparison between the GM and the free space map provided, black is the mapped occupied cells, and gray is the 2D projection of free-space map

or hollow object) such as barrier or sign-posts can not be correctly mapped using the proposed method. This is due mainly to the sensor model grid resolution.

2) Classification Results: The following two scenarios extracted from the complete sequence show the classification performance of the algorithm in different cases. The classification output is represented on the GM, in superposition with the mapping, using two colors. Green cells are considered moving whereas blue cells are considered static. The first column contains zoomed parts of the GM around the vehicle with the classified objects. In the second column, the view of the scene is given (the camera has been installed at the same elevation than the lidar). The detected objects are manually highlighted here using bounding boxes in order to make the link between the detected objects in the GM and their position in the image.

- Intersection
  This scenario is illustrated by Figure 12. The ego-vehicle is stopped at an intersection and two vehicles cross in front of it. Vehicles come from both right and left sides with a high radial speed. The car is correctly detected whereas the van is partially miss-classified.

- Pedestrian crossing the street
  This scenario is illustrated by Figure 13. The vehicle approaches an intersection and stops in order to let two pedestrians go across the road. Pedestrians are partially miss classified since they walk slowly.

One can notice that the field of view of the camera (˜70°) is small in comparison with the lidar one (˜140°). So, objects on the side are detected by the lidar before being visible on the image.

Moving vehicles are well detected. Some miss-detections have been observed in case of long vehicles, because their long size induces a spacial accumulation during a significant time.

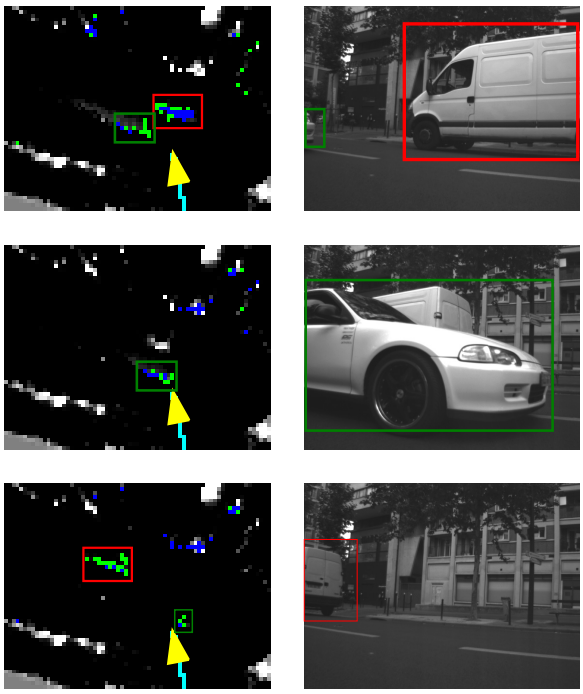

Fig. 12. Intersection scenario

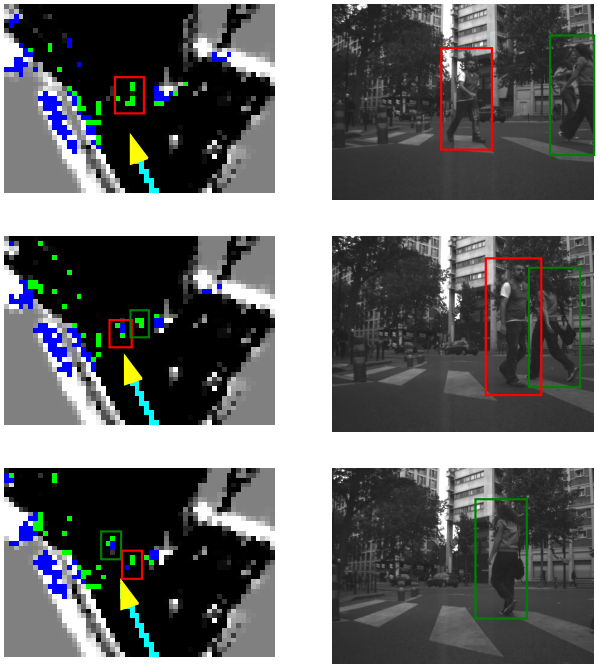

Fig. 13. Pedestrians crossing the street scenario

Other miss-detections occur if objects are moving too slowly like pedestrians waiting on the sidewalk before crossing the road. The classification results may be improve by using a higher resolution map.

V. Conclusion

This paper has presented a perception fusion scheme based on both local and global grid interactions. The contribution of this work is to propose a smart approach which can provide the navigation space, the mobile obstacles and static objects. In its current implementation, accurate localization is a prerequisite. The results we have reported here illustrate the nice performance of such a strategy for detecting mobile objects which is a crucial data processing step in perception. One perspective is to analyze how a dead-reckoning localization method, using for instance lidar odometry, will degrade the method. Another perspective is to use prior information like the IGN map in order to start with a prior map or to improve the detection performance by filtering object out of the road.

Here, the data fusion is simply done using positive and negative accumulation with saturation, the absence of knowledge being managed using no accumulation. This strategy has shown to be very efficient for capturing pieces of information coming from all the lidar's echoes. Finally, this fusion framework can be adapted to the implementation of other fusion paradigms such as the Bayesian or evidential ones. The evidential fusion method is our main perspective in order to handle the absence of information in a rigorous framework.

Acknowledgment

This work is supported by the French ANR CityVip project and the CityHome international ICT collaboration. The system has been tested on sequences recorded by the test platform "STEREOPOLIS". The authors would like to thank the CityVIP group and particularly the IGN staff.

References

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*, 2001.

[2] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46 – 57, 1989.

[3] Y. Zhao, H. Chiba, M. Shibasaki, R. Shao, X. Cui, and J.Zha, "Slam in a dynamic large outdoor environment using a laser scanner," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2008.

[4] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.

[5] Y. Bar-Shalom, *Multitarget-Multisensor tracking : Applications and Andvances*. Artech House, 2000.

[6] C.-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, September 2007.

[7] F. Fayad and V. Cherfaoui, "Tracking objects using a laser scanner in driving situation based on modeling target shape," *IEEE Intelligent Vehicles Symposium*, 2007.

[8] A. Petrovskaya and S. Thrun, "Model based vehicle detection and tracking for autonomous urban driving," *Auton. Robots*, vol. 26, no. 2-3, pp. 123–139, 2009.

[9] C. Coue, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessiere, "Bayesian occupancy filteing for multitarget tracking : an automotive application," *International Journal of robotics research*, vol. 25, no. 1, pp. 19–30, 2006.

[10] J. Borenstein, M. Ieee, Y. Koren, and S. M. Ieee, "Histogramic in-motion mapping for mobile robot obstacle avoidance," *IEEE Transactions on Robotics and Automation*, vol. 7, pp. 535–539, 1991.

[11] O. Garcia-Favrot and M. Parent, "Laser scanner based slam in real road and traffic environment," in *IEEE International Conference Robotics and Automation (ICRA09). Workshop on Safe navigation in open and dynamic environments Application to autonomous vehicles*, 2009.