# Grid based fusion of offboard cameras

Manuel Yguel, Olivier Aycard, David Raulo, Christian Laugier

# Grid based fusion of offboard cameras

Manuel Yguel, Olivier Aycard, David Raulo & Christian Laugier
GRAVIR-IMAG & INRIA RA
655 avenue de l'Europe - Montbonnot
38334 Saint Ismier Cedex - FRANCE
Email : Firstname.Lastname@imag.fr

*Abstract*— **The goal of a perception system is to built an environment model. This model could be a list of features and objects present in the environment but could also be a grid (ie, a discretization of the environment in cells), where each cell gives the probability that the corresponding part of the environment is occupied or not. In this paper, we describe the perception system, based on a grid environment model, developped in the frame of the French project PUVAME. This system consists of several offboard cameras observing an intersection to detect objects (ie, pedestrian, cyclists and vehicules...). We present a generic and new method to design a sensor model for offboard camera where each of the video camera feed is processed independantly by a dedicated detector. Moreover, to add tolerance to miss detections and false alarms, we model the failure of the sensor. We also detail how to build an occupancy grid, fusionning the information from the different cameras. Experimental results showing that our approach is well suited to build an environment model are provided.**

**Keywords:** perception, sensor data fusion, sensor model, occupancy grid

## I. Introduction

In France, about 33% of roads victims are VRU[1]. In its $3^{rd}$ framework, the french PREDIT[2] includes VRU Safety. The PUVAME project [1] was created to generate solutions to avoid collisions between VRU and Bus in urban traffic. An accident analysis has shown that an important part of these collisions take place at intersection and bus stop. To reduce accidents, a first requirement of the PUVAME project is to improve the perception of the bus driver at these particular places. This objective will be achieved using a combination of offboard cameras, observing intersections or bus stops, to detect VRU present at intersection or bus stop, as well as onboard sensors for localisation of the bus.

In this paper, we detail the solution we developped to fusion the information about the position of each VRU given by each offboard camera. The process of fusion is fundamental:

- As the measurement of every sensor always keeps a certain amount of uncertainty on the VRU position, it allows to use data coming from different sensors in order to compute a better estimation of the position of each object [12];
- It also increases the field of view of the whole perception system;
- Moreover, it is useful to decrease the level of false alarms.

In many applications, to perform fusion, a geometric point of view is used: a set of geometric features is first defined, a model of uncertainty associated to each feature is also needed and a way to fusion features has also to be provided. For instance, [2] used infrared camera and radar to detect and track road obstacle. Each sensor returns a point as observations of the position of each obstacle present in the environment. The uncertainty associated to this position is modelized by a gaussian and when two observations corresponds to the same obstacle a fusion of the two corresponding gaussian is performed to estimate the position of the object. In [4], a generalized feature model for the multi sensor case has been developped. This generalized feature model is based on the assumption that any entity in the world can be detected and recognized by means of features. Features are assumed to be dedicated parts of the entity with certain spatio-temporal coordinates in the coordinate system of the entity. Actually, the major drawback of the geometric approach is the number of different geometric features (points, segments, polygons, ellipses, etc) that the perception system must handle. Moreover, this approach is unable to take into account a new objet that appears in the environment and that could not be defined using the predefined set of features.

An other way to model the environment has been introduced by Elfes and Moravec at the end of the 1980s. This new framework to multi-sensor fusion is called Occupancy Grids (OG). An occupancy grid is a stochastic tesselated representation of spatial information that maintains probabilistic estimates of the occupancy state of each cell in a lattice [5]. The main advantage of this approach is the ability to integrate several sensors in the same framework taking the inherent uncertainty of each sensor reading into account, in opposite to the Geometric Paradigm whose method is to categorize the world features into a set of geometric primitives. The alternative that OGs offer is a regular sampling of the space occupancy, that is a very generic system of space representation when no knowledge about the shapes of the environment is available. The occupancy grid paradigm has been applied successfully in many different ways. For example, some systems use occupancy grids to plan collision-free paths [3] or for path planning and navigation [13] [7]. Therefore, most of actual mapping systems resort to OG for modelling the environment [13], [8]. And all the more so as with appropriate sensor models OG provide a rigorous way to manage occlusions in the sensor field of view. On the contrary of a feature based environment

[1]Vulnerable Road Users
[2]Programme de Recherche et d'Innovation dans les Transports Terrestres

model, the only requirement for an OG building is a bayesian sensor model for each cell of the grid and each sensor. This sensor model is the description of the probabilistic relation that links sensor measurement to space state, that OG necessitates to make the sensor integration. Fortunately it is possible for a wide class of sensors to factorise this amount of data by taking advantage of the characteristics of the sensor. Regarding telemetric sensors, sensor model for sonar [14] and laser range finders [13] have been defined and used to map the environment. 3D occupancy grids have been built using stereo vision [10] and a set of camera [6]. In these papers, the sensor model are defined using the result of a preprocessing of the images.

In this paper, we present a generic and new method to design a sensor model for a different kind of sensors: a set of offboard cameras where each of the video camera feed is processed independantly by a dedicated detector. The role of the detectors is to convert each incoming video frame to a set of bounding rectangles, one by target detected in the image plane. The set of rectangles detected at a given time constitutes the observation used to build the occupancy grid. This type of sensor model has never been defined before, and moreover we show that our approach is generic and could easily be adapted to any kind of visual detector. Moreover, to add tolerance to miss detections and false alarms, we model the failure of the sensor.

In next section, we present the experimental platform used to evaluate the solution we propose. Section III defines the occupancy grid basic concepts. In section IV and V, we detail how the sensor model of offboard cameras is built. Experimental results are reported in section VI. We give some conclusions and perspectives in section VII.

## II. ParkNav platform

The experimental setup used to evaluate our fusion scheme is an evolution of the ParkView platform, initially developped for a French national project designed for the Interpretation of Complex Dynamic Scenes and Reactive Motion Planning.

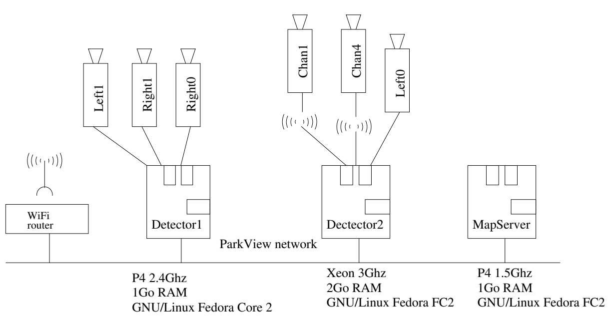
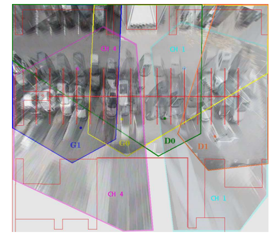



Fig. 1. The ParkView platform hardware

The ParkView platform is composed of a set of six off-board analog cameras, installed in a car-park setup such as their field-of-view partially overlap (see figure 2), and three Linux(tm) workstations in charge of the data processing, connected by a standard Local Area Network (figure 1).

(a)



(b)

(c) “left1”

(d) “left0”

(e) “right0”

(f) “right1”

Fig. 2. (a) Location of the cameras on the parking; (b) Field-of-view of the cameras projected on the ground; (c) to (d) View from four of the cameras

The workstations are running a specifically developped client-server software composed of three main parts, called the *map server*, the *map clients* and the *connectors* (figure 3).

The *map server* processes all the incoming observations, provided by the different clients, in order to maintain a global high-level representation of the environment; this is where the data fusion occur. A single instance of the server is running.

The *map clients* connect to the server and provide the users with a graphical representation of the environment; they can also process this data further and perform application-dependant tasks. For example, in a driving assistance application, the vehicle on-board computer will be running such a client specialized in estimating the collision risk.

The *connectors* are feeded with the raw sensor-data, perform the pre-processing, and send the resulting *observations* to the map server. Each of the computer connected with one or several sensors is running such a *connector*. For the application described here, all data preprocessing basically consist in detecting pedestrians. Therefore, the video stream of each camera is processed independantly by a dedicated detector. The role of the detectors is to convert each incoming video frame to a set of bounding rectangles, one by target detected in the image plane. The set of rectangles detected at a given time constitutes the detector observation, and is sent to the

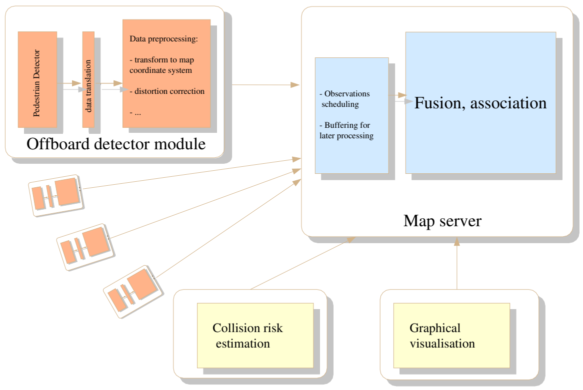



Fig. 3. The ParkView platform software oranization

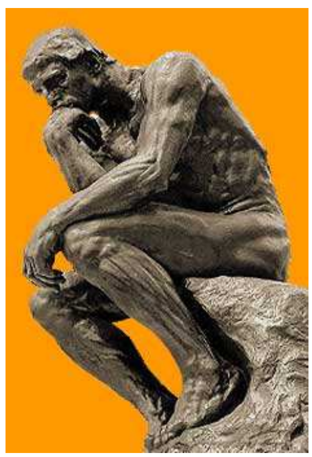


Fig. 4. A graphical map client

map server.

Since the fusion system operates in a fixed coordinate system, distinct from each of the camera's local systems, a coordinate transformation must be performed. For this purpose, each of the cameras has been calibrated beforehand. The result of this calibration consists in a set of parameters:

- the intrinsic parameters contain the information about the camera optics and CCD sensor: the focal length and focal axis, the distorsion parameters,
- the extrinsic parameters consist of the homography matrix: this is the 3x3 homogenous matrix which transform the coordinates of an image point to the ground coordinate system.

In such a multi-sensor system, special care must be taken of proper timestamping and synchronization of the observations. This is especially true in a networked environment, where the standard TCP/IP protocol would incur its own latencies.

The ParkView platform achieves the desired effect by using a specialized transfer protocol, building on the low-latency properties of UDP while guaranteeing in-order, synchronised delivery of the sensor observations to the server.

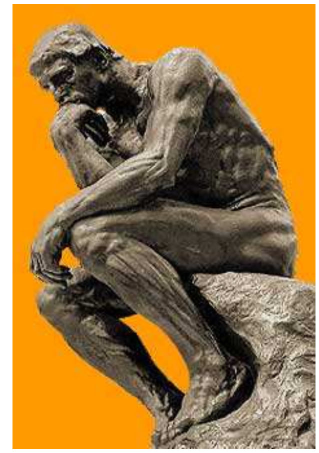


Fig. 5. The pedestrian detector

## III. Occupancy grid

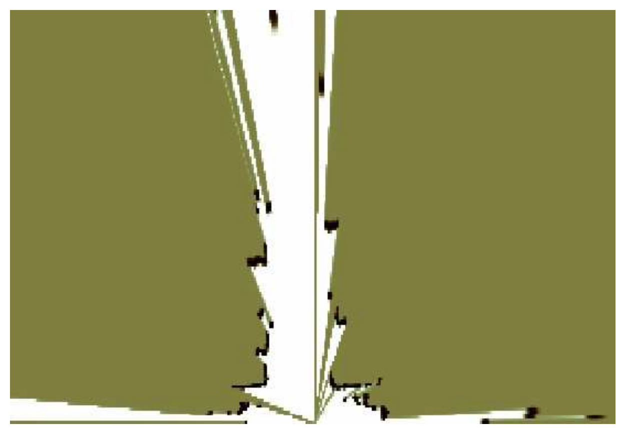


Fig. 6. Occupancy grid obtained from a laser range-finder.

### A. Definition

Elfes and Moravec have introduced at the end of the 1980s a new framework to multi-sensor fusion called Occupancy Grids(OG). An occupancy grid is a stochastic tesselated representation of spatial information that maintains probabilistic estimates of the occupancy state of each cell in a lattice [5], (Fig.6). It means that space is divided into cells in which is stored the information about the possibility that an obstacle lies in the cell or not. For each cell, observations can be made to modify the cell state. The observations come from sensors and the heart of the modelling problem is to define how each sensor measure modify the cell state.

*1) Mathematical Framework:* we introduce our framework and notation, deriving the update equations of a cell of the grid at each sensor measurement.

*a) Probabilistic variable definitions:*

- $\overrightarrow{Z} = (Z_1, \dots, Z_s)$ a vector of $s$ random variables, one variable for each sensor. We consider that each sensor

$i$ can return measurements from a set $\mathcal{Z}_i$ plus a special event "nothing measured" which means that the entire scanned region is free.

- $E_x \in \mathcal{E} \equiv \{\text{occ}, \text{emp}\}$. $E_x$ is the state of the bin $x$ either occupied ("occ") or empty ("emp"), where $x \in \mathcal{X}$.
  $\mathcal{X}$ is the set of indexes of all the cells in the monitored area.

For a certain variable $V$ we will note in capital case the variable, in normal case $v$ one of its realisation, and we will note $P(v)$ for $P([V = v])$ the probability of a realisation of the variable.

*b) Joint probabilistic distribution:* the lattice of cells is a type of markov field and many assumptions could be made about the dependencies between cells and especially adjacent cells in the lattice [9]. In this article we will explain sensor models for independant cells i.e. without any dependencies, which is a strong hypothesis but very efficient in practice since any calculus could be made for each cell apart. It leads to the following expression of a joint distribution for each cell.

$$P(E_x, \overrightarrow{Z}) = P(E_x) \prod_{i=1}^{s} P(Z_i | E_x) \tag{1}$$

Given a vector of sensor measurements $\overrightarrow{z} = (z_1, \ldots, z_s)$ we apply the bayes rule to derive the probability of cell $x$ to be occupied:

$$P(e_x | \overrightarrow{z}) = \frac{P(e_x) \prod_{i=1}^{s} P(z_i | e_x)}{P(\text{occ}) \prod_{i=1}^{s} P(z_i | \text{occ}) + P(\text{emp}) \prod_{i=1}^{s} P(z_i | \text{emp})} \tag{2}$$

For each sensor $i$, the two conditional distributions $P(Z_i|\text{occ})$ and $P(Z_i|\text{emp})$ must be specified. That what is called *the sensor model* definition.

### B. Sensor model

The definition of a bayesian sensor model is: the specification of a probability distribution over all the possible measurements of the sensor for each state of the cell: occupied and empty. What is very interesting with this definition is that it is purely functionnal, and so there are a lot of possiblities to acquire such a model. One can use sensor characteristics given by the sensor manufacturer or physical equations and it is also theoretically possible to learn the probabilistic distribution function with examples. We underline again that what is needed in occupancy grids design is a sensor model for each cell of the grid. Moreover it is possible for a wide class of sensors to factorise the amount of data of a sensor model per cell by taking advantage of the geometric symetries of the sensor and that is what we present in the next section for modelling the output of a video detector. This modelling is called sensor analysis and is particular for each sensor, moreover there is a part of this modelling: the fault model which follows the same design for each sensor.

*1) Fault modelling:* Thus, for each cell $x$ the function to define is: $P(Z_i|E_x)$, that is the conditional probability function of a sensor measurement knowing a cell state. It is possible to add a confidence model for each sensor so that it is possible to deal with the information about the amount of errors produced by a sensor. The principle is to consider a new variable:

- $D_i \in \mathcal{D} \equiv \{\text{on}, \text{off}\}$. $D_i$ is the state of the measurement, either correct ("on") or wrong ("off").

Now, the joint distribution to define is:

$$P(E_x, \overrightarrow{Z}, \overrightarrow{D}) = P(E_x) \prod_{i=1}^{s} P(F_i) P(Z_i | E_x, D_i) \tag{3}$$

that is defining $P(D_i)$ and defining $P(Z_i|Ex, \text{off})$ and $P(Z_i|Ex, \text{on})$. Defining $P(D_i)$ corresponds to define $P([D_i = \text{off}])$ which is simply the probability that the ith sensor produced a wrong measurement. $P(D_i|Ex)$ is assign to $P(D_i|Ex, \text{on})$ because it models the correct behaviour of the sensor. For $P(D_i|Ex, \text{off})$, without any kind of information, a non-informative distribution which assign the same probability to each sensor measurement, is chosen for the two possible states, $e_x$, of the cell.
If there is no information about the current behaviour of the sensor, the used distribution is just the marginalization over all the possible state of each measurement:

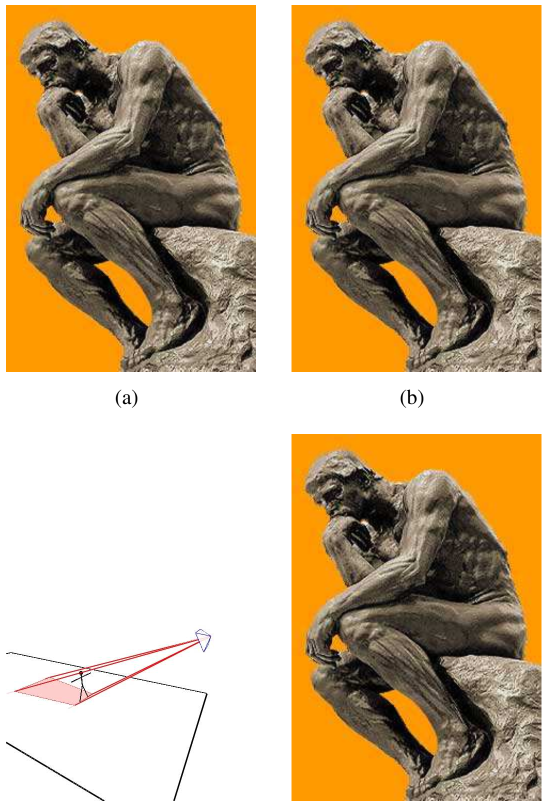
$$P(E_x, \overrightarrow{Z}) = P(E_x) \prod_{i=1}^{s} \sum_{\mathcal{D}} P(D_i) P(Z_i | E_x, D_i) \tag{4}$$

This kind of transformation of the sensor model adds a certain inertia related to the probability of wrong measurement. It means that a good sensor measurement must be seen $\frac{1}{P(\text{on})}$ times to be considered by the system as relevant as a sensor measurement without error model. This inertia is the price for the robustness added by the fault modelling.

### C. Building sensor models

In the two following sections we describe how building sensor models for a high level input, such as video camera motion detector output. The problem is that motion detectors give information in the image space and that we search to have knowledge in the ground plan. Thus suppose that the detector returns at most $n$ bounding boxes, mainly described by the pair of two points, the space of the possibles input values is $[1; M]^{4n}$ with an image of $M \times M$ pixels. Also without any simplification the problem is to build a probability distribution over a space of $M^{4n}$, for two reasonable values: $M = 256$ pixels and $n = 20$ the size is about $10^{192}$. So this problem must be converted into a more practical one. We propose there two different sensor models that are suitable for different purposes, but which underline the genericity of the occupancy grid approach. In both of the models we search first to segment the ground plan in three types of region: occupied, occulted and free zones using the bounding boxes informations. Then we introduce an uncertainty management to deal with the position errors in the detector. Finally, we explain how to convert this information into probability distributions and how to add fault model, controled by the quality estimation of the sensor output.

## IV. Sensor model of offboard camera with visibility of the ground-object contact points



(a) (b) (c) (d)

Fig. 7. (a) An image of a moving object acquired by one of the offboard video cameras and the associated bounding boxe found by the detector. (b) Schema of the viewing lines of the video camera that intersect the bounding boxe. (c) The occulted zone as the intersection of the viewing cone associated with the bounding boxe and the ground plan. (d) The associated ground image produce by the system.

### A. Image of the ground occupation

*a) One video camera, one bounding box:* the inputs of this environment modelling are output of video camera detectors that give bounding boxes of detected moving objects. A video camera only sees the visible surface of the objects in its field of view. Thus we have to draw on the ground the occupied, occluded and free zones. We make the following hypothesis:

1) the ground is a plan of which an equation is known.
2) all the obstacles stand on the ground: such as car, bycicles and non jumping pedestrians. This is statistically not a strong hypothesis in non flying objects tracking.
3) the part of the obstacles that is visible for the camera has an edge adjacent to the ground. This hypothesis is stronger because some times just a part of the obstacle is visible and the other part, which could be the bottom of the obstacle such as the legs for a pedestrian is occluded. But the higher are the video camera, the more true this hypothesis is, and in the case of roof camera with a field of view oriented toward the ground floor it is totally true when the second hypothesis holds.
4) all the surface of the returned bounding boxes is considered to hide the back of the scene.

According to hypothesis (2) and (3) the returned bounding box have a part adjacent to the ground. And this part is occupied by the obstacle. And according to hypothesis (1) and (4) the occluded zone by a bounding box is the projection of the bounding box to the ground according to the camera projection matrix (Fig. 7(b)).

That leads to the drawing of the ground occupation image with one bounding box. First we calculate the projection of the bounding box to the ground and we mark this area as occluded. It is possible because, the extrinsec parameters of the video camera have been calibrated before and the ground plan identified. Thus the projection of the bounding box is just the intersection of the cone of view defined by the bounding box with the ground plan (Fig. 7(c)). Second we calculate the part of the contour of the projected bounding box which is the closest to the video camera. We draw this contour with a certain width and mark it as an occupied area. We draw the rest as free (Fig. 7(d)).

*b) One video camera, several bounding boxes:* in the case of several bounding boxes, the projection of one can overlap the projection of an other. So that we have to handle carrefully the order of area drawing such as no occupied area will be marked as occluded or free when it was marked as occupied before. So we define three values: $\{0; 0.5; 1\}$ for free,occluded and occupied respectively. First we paint all the ground in free. Then we draw each bounding box with its occluded and occupied area and for each pixel the new value is just set to the max of the precedent value and the measured value.

### B. Position uncertainty

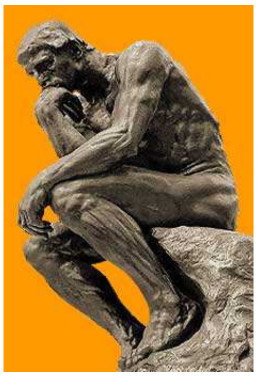
To handle position uncertainty due to video camera vibration, noise in the video detector, non perfect synchronisation of all the sensor measurements or the communication lattency we just make a convolution of the ground image obtained in the precedent step, with a gaussian 2D-kernel. The variances of these kernels are important parameter and in fact it suits the worst of the precedent sources of position uncertainties, Fig. 8(a).

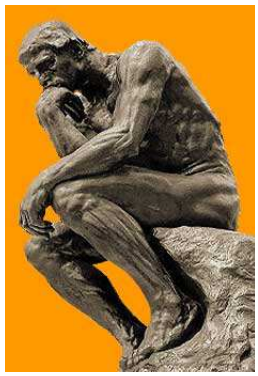
### C. Building the two maps of probabilities: $P(Z|Ex)$

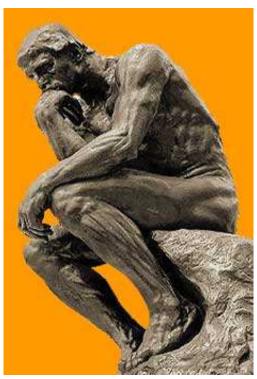
For each bin the precedent step provide a floating number $z \in [0; 1]$ describing the fact that there is or not an obstacle in the cell. A possible definition of the probability of this number for each possible state of the cell: $\text{emp}, \text{occ}$ is in term of probability density:

$$p(z|\text{emp}) = 2(1-z) \quad (5)$$
$$p(z|\text{occ}) = 2z \quad (6)$$

The main information is that the close $z$ is to 1, the most probable is the measure $z$, if the cell is occupied. For $P(Z|\text{occ})$ any increasing function over $[0;1]$ which integral is $1.0$ suits. Symetrically for $P(Z|\text{emp})$ any decreasing function over $[0;1]$ which integral is $1.0$ suits. We chose very simple functions: that are linear functions and reach $0$ and $1$ at their maxima.


(a)


(b)


(c)

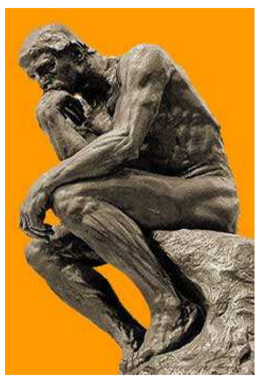

(d)

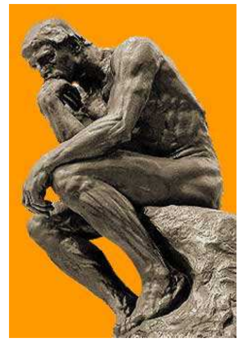
Fig. 8. (a) Ground image after gaussian convolution. (b) Probability of the ground image pixel value, knowing that the pixel corresponds to an empty cell: $P(Z|\text{emp})$ for each cell. (c) Probability of the ground image pixel value, knowing that the pixel corresponds to an occupied cell: $P(Z|\text{occ})$ for each cell. (d) The resulting probability that the cells are occupied after the inferrence process.

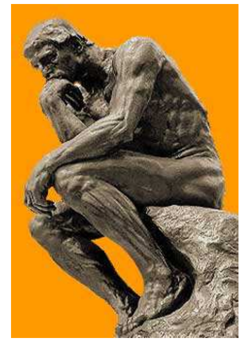
### D. Error model

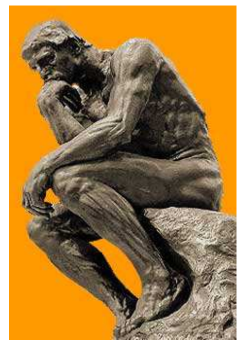
It is just adding constant images to the likelihood images obtained at the previous step weighted by the fault probability.

## V. General model of offboard camera with no visibility assumption

In this section we use the same inputs as in the previous one and in the process, the difference arises only in the drawing of the ground occupation image. Thus we focuses only on this part.


(a)


(b)


(c)

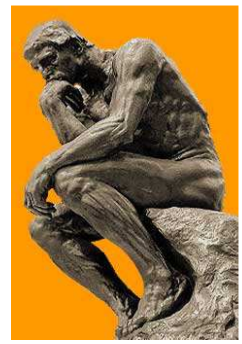

(d)

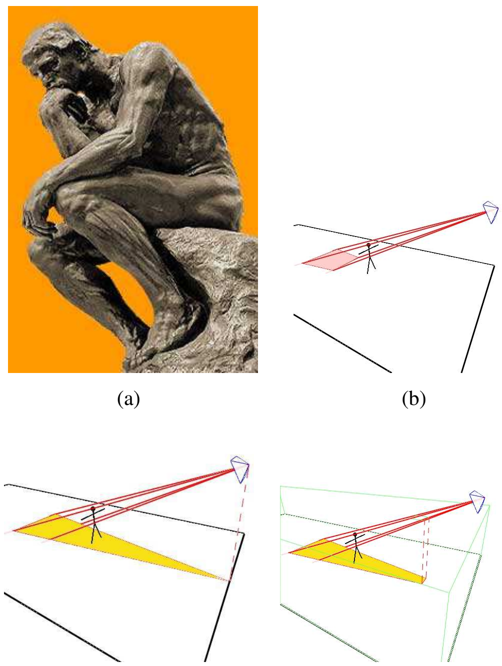
Fig. 9. (a) Probability of the ground image pixel value, knowing that the pixel corresponds to an empty cell: $P(Z|\text{emp})$ for each cell with an error model. (b) Probability of the ground image pixel value, knowing that the pixel corresponds to an occupied cell: $P(Z|\text{occ})$ for each cell with an error model. (c) The resulting probability that the cells are occupied after the inferrence process with an error model. (d) Difference of the probability that the cells are occupied after the inferrence process between the model without error and the model with error.

### A. Image of the ground occupation

*c) One video camera, one bounding box:* We consider, now, a more general sensor model, in the sens that the hypthesis (3) is considered to be sometimes false. Consider the case where the ground-object contact points are not visible (Fig. 10(b)) in such a case the previous sensor model gives a totally wrong ground occupation. The correct assumption in this case is to consider that the whole wiev cone that corresponds to the bounding box could be occupied (Fig 10(c)). It leads to the disapearing of occluded area, because it is not possible to make any distance distinction. Consequently the system is less accurate but sufficient enough in a context of multi-sensor fusion to bring informations. To reduce the size of the area which is considered as occupied, we add to the set $\{1, 2, 4\}$ a new assumption:

5) the tracked objects have a maximal height $h$.

In the context of a car park we fixe the maximal height to $3$ meters. To calculate the occupied area, we use the linearity of the cone projection. To each vertex $K$ of the bounding box we can associate a view line which is the line $EK$ that connect the camera focal point $E$ to the vertex $K$. For each of these view lines associated to a bounding box, we apply the same



(a) (b)

(c) (d)

Fig. 10. (a) Moving object whom the contact points with the ground are occulted. (b) The intersection of the viewing cone associated with the bounding boxe and the ground plan which is far from the position of the object. (c) Projection of the entire view cone on the ground in yellow. (d) Projection of the part of view cone that fits the object height hypothesis (in green).

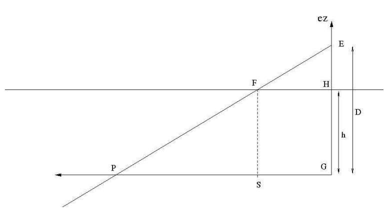



Fig. 11. Calculus of the position $S$ of front vertexes of the occupied area. They lie on the lines that join the orthogonal projection $G$ of the camera on the ground and the homography projection $P$ of the bottom of the bounding box on the ground. The distance $GS$ is obtained thanks to the Thales theorem (Eq. 7).

process. First we draw the projection $GP$ on the ground of the view line (Fig. 11, 10(c) red lines on the ground). Then we apply the Thales theorem (Fig. 11) line to calculate the position of the point $S$ that is the projection on the ground of $F$ the intersection of the view line with the plan $z = h$:

$$GS = HF \quad = \quad \frac{D-h}{D} * GP \qquad (7)$$

where $G$ and $H$ are the orthogonal projection of the camera focal point $E$ on the ground and on the plan $z = h$. $P$ is obtained by the homography projection of the bounding box point $K$ on the ground and $F$ on the plan $z = h$. For each view line we obtain 2 points $P$ and $S$. Thus for each bounding box we obtained $8$ points and as the bounding box is convex, so is its projection, then the resulting occupied area is the convex hull of these $8$ points.

In the particular case of a vertical bounding box in the image: the back points of the cone projection are obtained by the projection of the bounding box top points on the ground thanks to the homography matrix. Whereas the front points (Fig. 10(d)) are given by the projection (dotted lines) on the ground of the points that lies on the plan $z = h$.

The outside of the occupied area is drawn as free.

*d) One video camera, several bounding boxes:* In the case of several bounding boxes we apply the same algorithm as above, each pixel on the ground receives the max of the value that each bounding box assigns for it. In this particular case there is only two values: $\{0; 1\}$ empty or occupied.

## VI. Experimental results and discussion

### *A. Experiments with visibility of the ground-object contact points*

### *B. Experiments with no visibility assumption*

### *C. Interest of the occluded labelling of the zones*

### *D. Quality of position estimation*

### *E. Sensor model comparison*

The first model is precise, but only when it hypothesis holds. However we can assume that modern object detectors will be able to infer the position of the ground-object contact points even the entire object is not visible. In particular with pedestrians, we hope that the visibility of only a part of the body such as the head or the tronc will be sufficient enough to obtained a feet position estimation. In such cases this model will be the most suitable for position estimation. With the second model, the position uncertainty allows to surround the real position of the detected object, such that with other viewing points or other sensors, like laser range-finders or radar it is possible to obtained a good hull of the ground object occupation. Thanks to the uncertainty this last model will never give wrong information about the emptyness of an area, which is a garantee for safety applications.

## VII. Conclusion and perspectives

In this article we present an occupancy grid fusion framework with offboard video camera which show all the advantages of this kind of fusion in term of accuracy and adaptability. Accuracy because it leads to integrate occlusion informations which allow to make the observations coherent and increase the fiability of the whole fusion system. And adaptability because it is very easy to integrate different kinds of information processing as it is shown with the presentation, here, of two different sensor models. Because it was widely demonstrated that it suits particularly well any kind of telemetric sensor models as a low level data process, we promote this environment modelling as a base for a multi-objects tracking. We use the output of this stage of fusion process in order to extract new observations. This extraction process decrease widely the number of observations such that the next association stage becomes really fast. Then this article presents just a part of our multi-target tracking plateform, but the second part consist of a standard tracking system as described in [11].


## ACKNOWLEDGMENT

PUVAME is a project of the “Security oriented technologies” operational group of the PREDIT The partners of PUVAME thank the French PREDIT for supporting the work of this project.


## REFERENCES


[1] O. Aycard, A. Spalanzani, J. Burlet, T. Fraichard, C. Laugier, D. Raulo, and M. Yguel. Puvame - new french approach for vulnerable road users safety. In *IEEE International Conference on Intelligent Vehicules*, 2006. In submission.

[2] C. Blanc, L. Trassoudaine, Y. Le Guilloux, and R. Moreira. Track to track fusion method applied to road obstacle detection. In *International Conference on Information Fusion*, 2004.

[3] J. Borenstein and Y. Koren. The vector field histogram - fast obstacle avoidance for mobile robots. *IEEE Journal of Robotics and Automation*, 7(3):278–288, June 1991.

[4] H. Cramer, U. Scheunert, and G. Wanielik. Multi sensor fusion for object detection using generalized feature models. In *International Conference on Information Fusion*, 2003.

[5] Alberto Elfes. *Occupancy grids: a probabilistic framework for robot perception and navigation*. PhD thesis, Carnegie Mellon University, 1989.

[6] J. S. Franco and E. Boyer. Fusion of multi-view silhouette cues using a space occupancy grid. In *IEEE International Conference on Computer Vision*, 2005.

[7] D. Kortenkamp, R.P. Bonasso, and R. Murphy. *AI-based mobile robots*. MIT Press, 1998.

[8] B. J. Kuipers. The spatial semantic hierarchy. *Artificial Intelligence*, (119):191–233, 2000.

[9] Stan Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, 2001. Series: Computer Science Workbench, 2nd ed., 2001, XIX, 323 p. 99 illus., Softcover ISBN: 4-431-70309-8.

[10] H.P. Moravec and M.C. Martin. Robot navigation by 3d spatial evidence grids. Technical report, Mobile Robots Laboratory, Robotics Institute, Carnegie Mellon University, 1994.

[11] D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. People tracking with a mobile robot using sample-based joint probabilistic data association filters. *International Journal of Robotics Research (IJRR)*, 2003.

[12] B. Steux, C. Laurgeau, L. Salesse, and D. Wautier. A vehicule detection and tracking system featuring monocular color vision and radar data fusion. In *IEEE International Conference on Intelligent Vehicules*, 2002.

[13] S. Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 1(99):21–71, 1999.

[14] B. Yamauchi. A frontier based approach for autonomous exploration. In *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pages 146–151, 1997.