Author manuscript, published in "IEEE Transactions on Intelligent Transportation Systems (2012)"
DOI : 10.1109/TITS.2012.2188393

A visibility-based approach for occupancy grid computation in disparity space

Mathias Perrollaz*, John-David Yoder†, Amaury Nègre‡, Anne Spalanzani*, Christian Laugier*
*Inria Grenoble - Rhône-Alpes, mathias.perrollaz@inria.fr
†Ohio Northern University, j-yoder@onu.edu
‡CNRS, Grenoble Informatics Laboratory

hal-00671208, version 1 - 16 Feb 2012

***Abstract*—Occupancy grids are a very convenient tool for environment representation in robotics. This paper will detail a novel approach to compute occupancy grids from stereo-vision, and shows its application for the field of intelligent vehicles. In the proposed approach, occupancy is initially computed directly in the stereoscopic sensor's disparity space. The calculation formally accounts for the detection of obstacles and road pixels in disparity space, as well as partial occlusions in the scene. In a second stage, this disparity-space occupancy grid is transformed into a Cartesian space occupancy grid to be used by subsequent applications. This transformation includes spatial and temporal filtering. The proposed method is designed to be easily processed in parallel. Consequently, we chose to implement it on GPU, which allows real-time processing for the demanding application. In this paper, we present this method and we propose an application to the problem of perception in a road environment. Results are presented with real road data, comparing qualitatively this approach with others.**

I. Introduction

Occupancy grids have been used for a variety of applications in the field of robotics, particularly for mapping the environment. Recent approaches for dynamic grid processing allow efficient use of occupancy grids for the monitoring of highly dynamic scenes, so that grids become a great tool for generic obstacle detection. Occupancy grids have typically been created based on data provided by range sensors such as laser or ultrasound. Current practice is to create the grids based on a probabilistic sensor model such as [1]. The use of stereo-vision to create occupancy grids is somewhat less common, but has recently raised a great interest in the community.

This paper presents a novel approach for the construction of occupancy grids using a stereo camera pair, which has been specifically created for the application of an on-road intelligent vehicle - as such, the occupancy grid created will be a plane representing the area in front of the vehicle. The method presented here provides a formal probabilistic model to calculate the probability of occupancy based on the disparity space of the stereoscopic sensor. The method formally considers the geometrical visibility of the different regions of the image in the calculation, and thus can deal with partially-occluded objects. It also accounts for the possibility of the road and obstacles being visible in the same cell of a grid. Because of its usage of the disparity space, it is computationally efficient, especially on a parallel architecture. It also considers the reduction in accuracy of the disparity measurements with distance from the sensor. After occupancy grid mapping, the Bayesian Occupancy Filter (BOF) framework is used to filter the grid over time. The paper will detail the methodology, and show results with real road data.

Section II provides a brief review of related work, dealing with stereo-vision based environment modeling and grid mapping. Section III gives some reminders on the use of the disparity data, with specifics related to the intelligent vehicle application. Section IV explains the methodology used for building an occupancy grid in this disparity space. Section V presents how this occupancy grid is transformed and filtered to obtain a usable occupancy grid in the Cartesian space. Section VI briefly describes the implementation, shows results, and compares the method with other approaches. Finally, section VII concludes and discusses future work.

II. Related work

In the field of mobile robotics, occupancy grids [2] are frequently used for modeling the environment. In this representation, each cell of a grid represents a region of the environment and is associated to a probability of the region being occupied by an object. The occupancy grid framework is popular for mapping static environments, and recent advances have improved this framework by adding the ability to represent dynamic environments. For this purpose, the Bayesian Occupancy Filter (BOF) [3] [4] combines the occupancy grids with velocity grids, and implements the Bayesian filtering methodology in this framework. An extension of this approach, which adds prior knowledge from a map, is proposed in [5]. A particle-based approach to such dynamic grids is also presented in [6].

Occupancy grids are typically constructed from beam-type range sensors, such as laser [1] or ultrasound [2]. Stereo camera pairs have been less commonly used for the creation of occupancy grids, because of the necessary processing time and the limited accuracy. Stereo-vision is more often considered for obstacle detection [7] [8] [9]. In this case, the idea is to convert the stereo data into an object-based representation. Two main categories of algorithms can be considered for that. The first category works in Cartesian

hal-00671208, version 1 - 16 Feb 2012

space: the disparity data is first converted into a large point cloud of 3D data, and then used for processing. In this category, many methods have been proposed. For example, in [10], the authors project this cloud vertically on a plane and apply a threshold for detection. In [11], the authors find positions where a fixed-size volume contains a sufficient number of points. In [12] the point cloud is clustered through the computation of histograms. Typically with Cartesian-based approaches, the problem of segmentation is not very different between using a laser or vision, particularly for short range where the data can be very precise. For medium to long range, data becomes very sparse, due to the sampling over integer pixel and disparity values. The sparse data causes the connectivity of points from the same object to be lost and algorithms based on proximity may fail. Before this aggregation step the authors in [9] propose to re-sample these points to ensure that their density is independent of the range. To avoid the problem of non-constant accuracy of stereo-vision, the second category of approaches focuses on working directly in the disparity space. This is the case for the u-v-disparity approach [7] or for approaches based on connectivity [13].

Similar to these object-based approaches, both 3D-based and disparity-based approaches exist for building occupancy grids using stereo-vision. With 3D-based approaches, the stereo cameras are used purely as a distance sensor, with a similar approach to a laser sensor [14]. Some authors have used more stereo-specific methods, but have not completely considered the nature of the stereoscopic data. In [15] or [16], the authors only consider the first detected object for each column and suppose that it occludes the field ahead. The first object is obtained by finding the highest disparity value for each column of the image. The result is a ternary grid (free/occupied/unknown) and the approach is sensitive to outliers in the disparity map. A great improvement has been proposed in [17], where the authors propose to build such a grid in the u-disparity plane. Occupancy is then robustly estimated using dynamic programming. The result is a ternary grid, specifically designed for free field estimation. This approach is used in [18] to build a medium level decomposition, called stixels, combining the advantages of disparity approach (pixels, connectivity) and 3D approaches (objects dimensions) in a compact representation.

The method proposed in this paper differs from the approaches outlined above in that it works in the u-disparity plane and provides a probabilistic grid (not ternary). It also takes advantage of some of the capabilities of a visual sensor, notably the ability to find partially occluded objects, and the ability to detect the road.

III. The data in the disparity space

A. Geometrical considerations

In this paper the stereoscopic sensor is considered as perfectly rectified. Cameras are supposed identical and classically represented by a pinhole model, $(\alpha_u, \alpha_v, u_0, v_0)$ being the intrinsic parameters. Pixel coordinates in left and right cameras are respectively named (u_l, v) and (u_r, v). The length of the stereo baseline is b_s.

A world coordinate system is denoted R_w. Each point P of coordinates $X_w = (x_w, y_w, z_w)$ can be projected onto the left and right image planes, respectively on positions (u_l, v) and (u_r, v). Consequently, in the disparity space associated to the stereoscopic sensor, the coordinates of P are $U = (u, d, v)$, with $u = u_l$ and $d = u_l - u_r$, namely the disparity value of the pixel. The u, d and v axes define the disparity coordinate system R_Δ. The transform $U = F(X_w)$ is invertible, so the coordinates in R_w can be retrieved from images coordinates through a reconstruction function.

For simplicity in notation and without loss of generality (see section III-B4), the yaw, pitch and roll angles of the camera, relative to R_w, are set to zero. Assuming that the center of the stereo baseline is situated at position (x^o, y^o, z^o), the transform from the world coordinate system to disparity space is given by [19]:

$$\begin{cases} u &= u_0 + \alpha_u \frac{x_w - x^o - b_s/2}{y_w - y^o} \\ v &= v_0 + \alpha_v \frac{z_w - z^o}{y_w - y^o} \\ d &= \alpha_u \frac{b_s}{y_w - y^o} \end{cases} \quad (1)$$

B. The u-disparity approach

1) The idea: The u-disparity approach [20] is a complement to the v-disparity originally described in [7]. The idea is to project the pixels of the disparity map along the columns, with accumulation. The resulting image is similar to a bird's eye view representation of the scene in the disparity space.

2) The detection plane: For occupancy grid computation, we have to consider a detection plane $\mathcal{P}_D$, which is the support for the grid. $\mathcal{P}_D$ is chosen to be parallel to the plane defined by the baseline and the optical axes. A coordinate system $R_D(O_D, \vec{x_d}, \vec{y_d})$ is associated with the detection plane. For a given point P of the space, $x_d = x_w$ and $y_d = y_w$. Arbitrarily, one can decide to set the detection plane to $z_w = 0$.

Considering equation 1, it can be seen that an orthogonal projection on $\mathcal{P}_D$ is equivalent to an orthogonal projection in R_Δ on any plane of constant v. Therefore, since computation of u-disparity images is not costly, this approach directly implements the vertical projection on $\mathcal{P}_D$ of the observed points from the scene. Moreover, it is equivalent to process the data in the u-disparity plane or in $\mathcal{P}_D$. For the reminder of this paper, we will call U_D the coordinates of a point in the u-disparity plan and X_D its coordinates in the detection plane. The transform between U_D and X_D is given by the reconstruction function G_D:

$$\begin{aligned} G_D : \quad & \mathbb{R}^2 \to \mathbb{R}^2 \\ & U_D \mapsto X_D \end{aligned} \quad (2)$$

with:

$$\begin{cases} x_d &= x^o + \frac{b}{2} + \frac{b(u - u_o)}{d} \\ y_d &= y^o + \frac{\alpha_u b_s}{d} \end{cases} \quad (3)$$

hal-00671208, version 1 - 16 Feb 2012

3) The alignment of rays: A major reason for calculating the occupancy in u-disparity space is that it allows us to represent all the optical directions of light rays passing through the camera's sensor as parallel (something that cannot be done on a discretized Cartesian grid). Then a set of vertically aligned rays is represented by a column in the u-disparity image. This provides major advantages for our approach, notably allowing for quick calculation of the visibility and occlusion of the image at varying distances from the camera.

4) Alternative detection planes: The proposed definition of the detection plane is not restrictive. If the stereoscopic sensor cannot be mounted parallel to the detection plane, an homography can be applied to the images in order to retrieve the expected geometrical configuration. Readers interested in this classical computer vision problem can refer to [21].

C. Road-obstacle separation

Critical to the performance of the method described below is the assumption that pixels appearing in the disparity image can be distinguished as being from the road surface or from the obstacles. This has been accomplished in several ways, such as estimating the road surface and thresholding the height of the pixels [14]. We choose to use a double correlation framework, which exploits different matching hypotheses for vertical and horizontal objects, as described in [22] and detailed in [19]. It provides immediate classification of the pixels during the matching process, without defining any threshold. After this classification, we obtain two disparity images I_d^{obst} and I_d^{road} and two u-disparity images, I_U^{obst} and I_U^{road}, respectively containing pixels from the obstacles and from the road surface.

IV. Occupancy grid computation in the disparity space

A. The approach

The approach presented here is to compute an occupancy grid directly in the u-disparity plane. This grid will later be transformed into a Cartesian grid, explicitly modeling the uncertainty of the stereoscopic sensor. There are two main advantages to this approach. First, it allows estimation of the visibility of each portion of the disparity space. Second, it allows the use of equally-spaced measurement points to create the initial grid. By contrast, moving to a Cartesian space first would give a varying density of measurements.

B. Notations

In this paper, we will use binary random variables. Considering such a variable B, we will write $P(B)$ and $P(\neg B)$ the probability density functions respectively associated to the hypotheses $B = 1$ and $B = 0$.
We will also use the notation U_D to represent either a pixel of coordinates (u,d) in the u-disparity plan or a cell in the u-disparity grid (which has pixel-wise resolution).

For each cell U_D of the grid, let us define a set of binary random variables:

- V_U: visibility of the cell. $V_U = 1$ means that the cell is visible,
- C_U: "obstacle confidence" of the cell. $C_U = 1$ means that an obstacle is seen in the cell,
- R_U: "road confidence" of the cell. $R_U = 1$ means that the road surface is seen in the cell,
- O_U: occupancy of the cell, based on obstacle pixels. $O_U = 1$ means that the cell is occupied by obstacle pixels,
- T_U: total occupancy of the cell. $T_U = 1$ means that the cell is occupied, according to both obstacle and road pixels.

C. Estimation of the occupancy of a cell, using obstacle pixels

We seek to calculate $P(O_U)$, the probability that a cell U_D is occupied, based on the obstacle disparity map. This probability will depend on the visibility, V_U, and on the "obstacle confidence" C_U. In order to estimate the shape of the probability density function $P(O_U|V_U, C_U)$, that is, the probability of a cell being occupied knowing V_U and C_U, some boundary conditions of $P(O_U|V_U, C_U)$ are fixed intuitively. For example, if the cell is not visible, nothing is known about its occupancy, so:

$$\forall c \in \{0, 1\}, P(O_U|\neg V_U, C_U = c) = 0.5 \tag{4}$$

Similarly, if a cell is fully visible and there is full confidence that an obstacle was observed, then:

$$P(O_U|V_U, C_U) = 1 - P_{FP} \tag{5}$$

that is, the only way the cell is not occupied is in the event of a false positive. Also:

$$P(O_U|V_U, \neg C_U) = P_{FN} \tag{6}$$

that is, a cell can only be occupied, when nothing is observed, if there was a false negative. P_{FP} and P_{FN} are respectively the probability that a false positive or a false negative can occur during the matching process. These are assumed to be constant and known.
Finally, the laws of probability are used to obtain the full decomposition of $P(O_U)$:

$$P(O_U) = \sum_{v,c} P(V_U = v)P(C_U = c)P(O_U|V_U = v, C_U = c) \tag{7}$$

V_U and C_U being boolean variables, this means:

$$\begin{aligned} P(O_U) = \quad & P(V_U)P(C_U)(1 - P_{FP}) \\ & + P(V_U)(1 - P(C_U))P_{FN} \\ & + (1 - P(V_U)) \cdot 0.5 \end{aligned} \tag{8}$$

So to compute the occupancy probability, it is necessary to estimate the values $P(V_U)$ and $P(C_U)$ with respect to the disparity data.

hal-00671208, version 1 - 16 Feb 2012

1) Classification of the pixels: For a given cell $U_D = (u, d)$ of the grid, we propose to define three sets of pixels in I_d^{obst}: possible pixels, visible pixels and observed pixels.
Possible pixels for U_D are defined as:

$$\mathrm{S}_P(U_D) = \{(u, v), v \in [v_h(d), v_0(d)]\} \tag{9}$$

$v_0(d)$ and $v_h(d)$ being respectively the v-coordinates of the pixels situated on the ground ($z_w = 0$) and at the maximum detection height ($z_w = -h$) for the value d of the disparity. Then the cardinality of $\mathrm{S}_P(U_D)$ is simply:

$$N_P(U_D) = v_h(d) - v_0(d) \tag{10}$$

Visible and observed pixels are obtained by classifying the possible pixels of U_D according to the following heuristic:
For a pixel $P(u, v) \in \mathrm{S}_P(U_D)$,

- if $I_d^{obst}(u, v) > d$, P is occluded,
- if $I_d^{obst}(u, v) = 0$, there is no observation for the ray (u, v) (*i.e.* P is not visible),
- else P is said to be visible. In this case, if $I_d^{obst}(u, v) = d$, P is an observed pixel for U_D.

This allows measurement of the number of observed pixels $N_O(U_D)$ and visible pixels $N_V(U_D)$ for the cell U_D.

2) Estimation of the visibility of a cell: The probability of visibility is estimated to be the ratio between visible pixels and possible pixels:

$$P(V_U) = \frac{N_V(U_D)}{N_P(U_D)} \tag{11}$$

3) Estimation of the confidence of observation: We choose to express the "obstacle confidence" as a function of the ratio:

$$r_O(U_D) = \frac{N_O(U_D)}{N_V(U_D)} \tag{12}$$

This means that if more of the visible pixels are filled with an observation, we are more confident we have observed an obstacle. An exponential function is used to represent the knowledge that the confidence should grow quickly with respect to the number of observed pixels:

$$P(C_U) = 1 - e^{-\frac{r_O(U_D)}{\tau_O}} \tag{13}$$

where τ_O is a constant parameter.

4) Resulting probability density function: Figure 1 illustrates equation 8, the probability density function of occupancy with respect to the visibility and to r_O, for $\tau_O = 0.1$. Knowing this function, the computation of the obstacle occupancy grid in the u-disparity image is straightforward.

D. Improving the occupancy estimation, using road pixels

As stated earlier, part of the matching involves separating the road pixels from the obstacle pixels in the disparity image. The prior sections described finding the probability of occupancy by an obstacle. However, the quality of the occupancy grid can be improved based on the use of the road pixels. Call $P(T_U)$ the total occupancy probability for cell U_D, considering both road and obstacle pixels, and call R_U

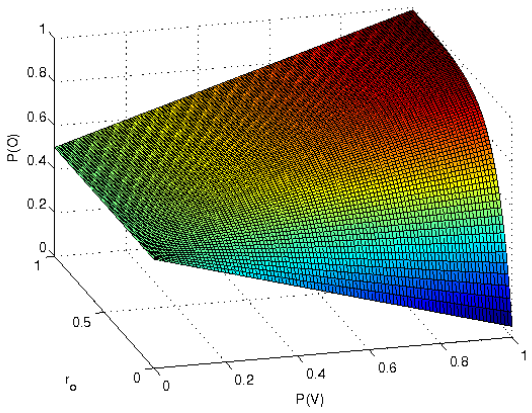


Fig. 1. *Probability density function* $P(O_U)$.

the binary random variable meaning that cell U_D only belongs to the road surface. We begin with the logical assertion that the cell is totally occupied if it is occupied by an actual obstacle and not by the road surface:

$$P(T_U) = P(O_U) \cdot P(\neg R_U) = P(O_U) \cdot (1 - P(R_U)) \tag{14}$$

To compute $P(R_U)$, we consider both obstacle and road pixels. This is because road pixels are often found at the base of obstacles, meaning that $P(R_U)$ must remain low when $P(O_U)$ is high.
Contrary to obstacle pixels, road pixels do no accumulate much over the u-disparity plane (they are not vertically aligned). Therefore, instead of using the accumulation value $I_U^{road}(U_D)$ we prefer measuring the number of road pixels in the neighborhood of U_D. Let us call $r_R(U_D)$ the ratio of non-zero road pixels in the $3 * 3$ neighborhood of U_D. Note that this value can be efficiently computed using a basic image filtering operation, with an all-one $3 * 3$ convolution kernel. We propose to compute $P(R_U)$ as:

$$P(R_U) = e^{-\frac{1-r_R(U_D)}{\tau_R}} \cdot e^{-\frac{r_O(U_D)}{\tau_O}} \tag{15}$$

τ_R being a constant parameter. Figure 2 shows the resulting probability density function for $P(R_U)$, with $\tau_R = 0.1$.

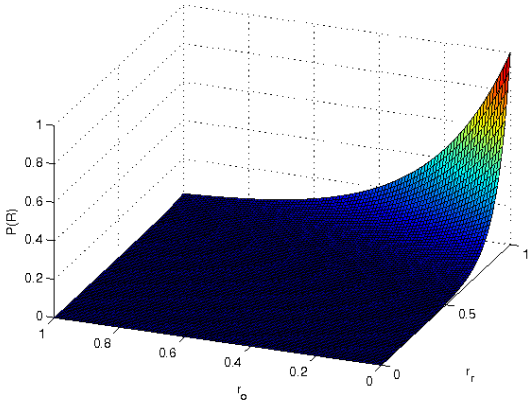


Fig. 2. *Probability density function* $P(R_U)$.

hal-00671208, version 1 - 16 Feb 2012

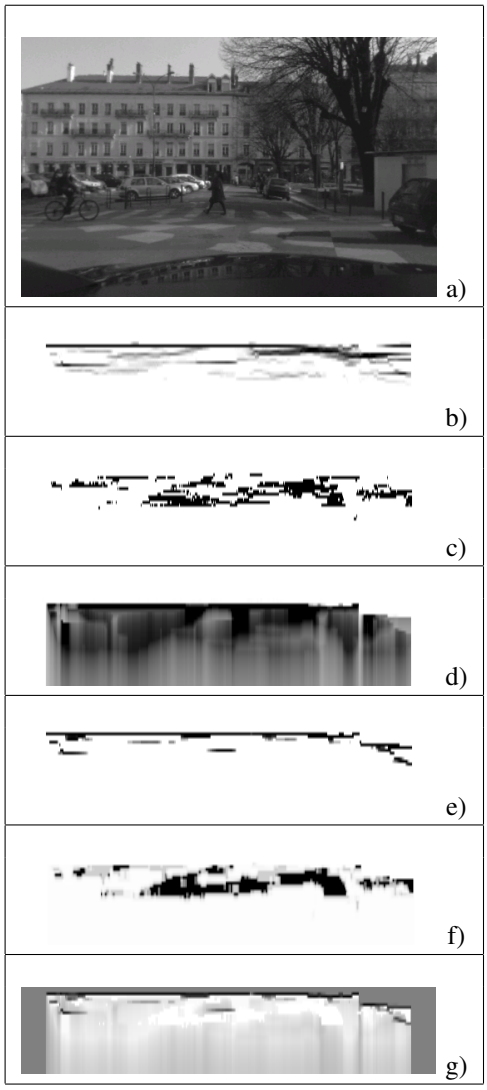


Fig. 3. *Computation of the occupancy grid in the u-disparity plane: a) left image from the stereo pair; b) obstacle u-disparity image; c) road u-disparity image; d) visibility map (black represents visible parts); e) "obstacle confidence" map (black represents high confidence); f) "road confidence" map (black represents high confidence); g) resulting occupancy grid in the u-disparity plane (black is occupied, white is free).*

E. Resulting occupancy grid

Figure 3 shows the basic application of this algorithm. a) shows the image from the left camera. b) and c) show the separated u-disparity images I_U^{obst} and I_U^{road}; d) e) and f) respectively show the "visibility", "obstacle confidence" and "road confidence" maps; and g) shows the u-disparity occupancy grid. In b), what you see are the fronts of obstacles, resulting in (mostly) straight black lines. Note that in the u-disparity images, you get closer to the camera as you get lower in the image, so the lower obstacle in the center of the image is corresponding to the pedestrian on the center, the upper line corresponds to the buildings, etc. It is also worth noting that the relationship between distance in the u-disparity image and Cartesian space is nonlinear, since pixels further from the camera represent larger areas. The road u-disparity image (c), meanwhile, shows much more detail where there is dense information on the road, such as on the crosswalk. You can see that in the occupancy grid g) black pixels represent high probability of occupancy, and white ones very low probability. This occupancy grid maintains strong information from the obstacles (they remain black lines), while a cell is empty (white) in areas where the road was detected. Most areas behind obstacles are completely unknown (gray), meaning they are assigned a value of $P(T_U) = 0.5$. Note however that partially-occluded objects (such as the building behind the pedestrian) are still found.

V. Computation of the occupancy grid in Cartesian space

A. Inverse mapping

The Cartesian occupancy grid is obtained by remapping the u-disparity grid over the detection plane, similar to stereo inverse perspective mapping [23]. This requires the calculation of which pixels from the occupancy grid in u-disparity have an influence on a given cell of the Cartesian grid. Let us define the surface $\mathrm{Surf}_U(U_D)$ of a pixel $U_D = (u, d)$ as the region of the u-disparity image delimited by the intervals $[u-0.5, u+0.5[$ and $[d-0.5, d+0.5[$. The area of influence of this pixel in the detection plane is: $\mathrm{Surf}_X(U_D) = G(\mathrm{S}_U(U_D))$. To compute the occupancy grid, the occupancy probability of a pixel U_D is simply attributed to the area $\mathrm{Surf}_X(U_D)$ of the detection plane.

For short distances, several pixels can have an influence on the same cell of the metric grid. It is necessary to estimate the occupancy according to this set of data. For this purpose, we choose to use a max estimator, which ensures a conservative estimation of the probability of occupancy:

$$P(T_X) = max(P(T_U)/X \in \mathrm{Surf}_X(U_D)) \qquad (16)$$

B. Filtering the occupancy grid

1) Spatial filtering: The occupancy grid obtained from the proposed method presents strong discretization effects, due to the pixel-level sampling and to the estimation of the disparity on integer values. We propose to apply a filter to this occupancy grid, in order to obtain a smoother and more realistic representation. To filter the occupancy grid an image-like filter, based on the convolution with a Gaussian kernel, is used. The problem is that a constant convolution kernel for the complete grid cannot be used, since the uncertainty of measurement is related to the range. Indeed, considering a constant kernel would require a strong trade-off when choosing the standard deviation of the filter. A small value would be accurate for close cells (low value of y_d) but inaccurate for distant cell, while a large value would suppress many details in the near environment.

Considering this fact, it is better to consider a constant Gaussian kernel in the u-disparity plane. The standard deviation σ_d^2 along the d axis is related to the disparity discretization (e.g. $\sigma_d = 0.5$). It can be relevant to consider that the standard deviation σ_u^2 along the u axis is related to the width of the correlation window to model effects like foreground fattening, as presented in [19].

hal-00671208, version 1 - 16 Feb 2012

2) Temporal filtering: The Bayesian Occupancy Filter (BOF) framework provides filtering capability, as well as the ability to estimate a velocity distribution for each cell of the grid. The BOF [3] is an adaptation of the Bayesian filtering methodology to the occupancy grid framework. It is based on a prediction-estimation paradigm. As an input, it uses an observed occupancy grid. On its output, it provides an estimated occupancy grid and also velocity grids, representing the probability distribution over possible velocities for each cell. In our case, we use the BOF for taking advantage of the time consistency among successive frames. This strategy allows removal of errors dues to short-lived matching errors. It also reinforces the estimated occupancy probabilities after several observations. An example of filtering with the BOF is displayed on figure 4

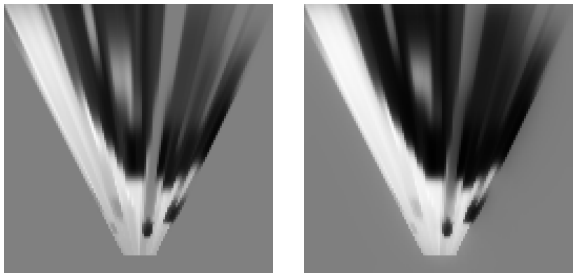


Fig. 4. *Effect of the BOF for filtering a Cartesian Grid: the grid on the right is filtered. It appears that the temporal filtering reinforces both the confidence on the free space on the left and the confidence on the occupied area on the right.*

VI. Experimental results

A. Experimental setup

Our experimental platform is a Lexus LS600h equipped with a TYZX stereo camera placed behind the windshield. The stereo baseline is $22cm$, with a field of view of $62°$. Camera resolution is 512×320 pixels with a focal length of 410 pixels. The on-board computer is equipped with 8GB of RAM, an Intel Xeon 3.4GHz processor and an NVIDIA GeForce GTX 480 for GPU.

The observed region is $20m$ long by $20m$ wide, with a maximum height of $1.8m$. Cell size for the occupancy grids is $0.2 \times 0.2m^2$. The correlation window used for matching measures 21pixels in width and in height. The occupancy grid computation parameters are set to: $P_{FP} = 0.02$ and $P_{FN} = 0.02$, $\tau_O = 0.1$ and $\tau_R = 0.1$.
The vehicle is also equipped with two IBEO Lux laser scanners. Each one provides four layers of up to 200 impacts, with a field of view of 100 degres.



Fig. 5. The experimental vehicle is a Lexus LS600h, equipped with a TYZX stereo camera, and two IBEO Lux laser scanners.

B. Results

U-disparity grids: Figure 6 compares our approach with a non-probabilistic approach on a simple example. The method depicted on figure 6-c) detects the maximum disparity of each column, as in [16]. Then, it provides a ternary grid, similar to the grids build from a 2D laser scanner (here, the occupancy value is P_{FN} for higher disparity values, and 0.5 behind the detected object). Our approach is displayed on figure 6-d).

The main interest of the visibility approach is seen on the pedestrian: it can be detected even with the box partially occluding its body. This can be convenient for fusion with a single layer laser scanner or for detecting objects faster.

Another difference is that our approach provides a probabilistic grid, with a more realistic variation in the occupancy probability values. A major interest of this aspect is that it is less sensitive to noise in the data (here matching errors). On the right of the box, such noise appears, causing the ternary method to fail. The probabilistic nature of our approach also permits to use it within probabilistic frameworks, like the BOF in our case.

Finally, it is also noticeable that our method uses additional information from the detection of the road surface, building a more precise description of the environment.

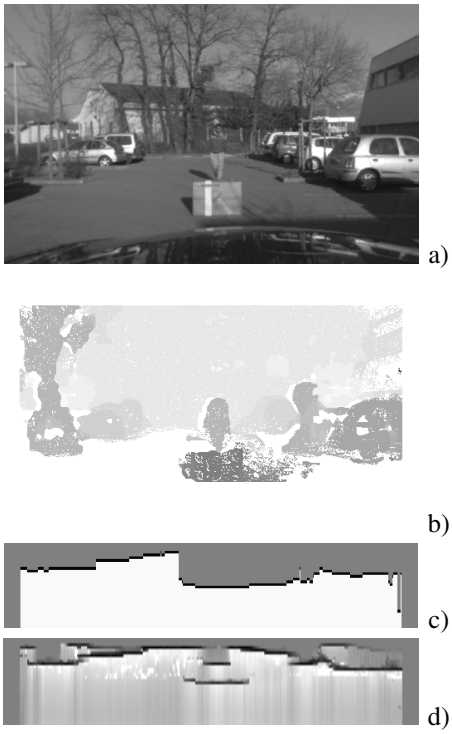


Fig. 6. *Comparison of our approach for computing the occupancy grid, with an approach which computes a ternary grid. a) left image from the stereo pair, b) corresponding obstacle disparity image, c) detection of the highest disparity value of each column, d) computation with the visibility approach.*

Figure 7 shows example results on more realistic driving scenes. For each case: the upper image is the left image from the stereo camera. For convenience in reading the results, and as an illustration of the potential behavior of our approach for detection, the regions of the image with an occupancy

hal-00671208, version 1 - 16 Feb 2012

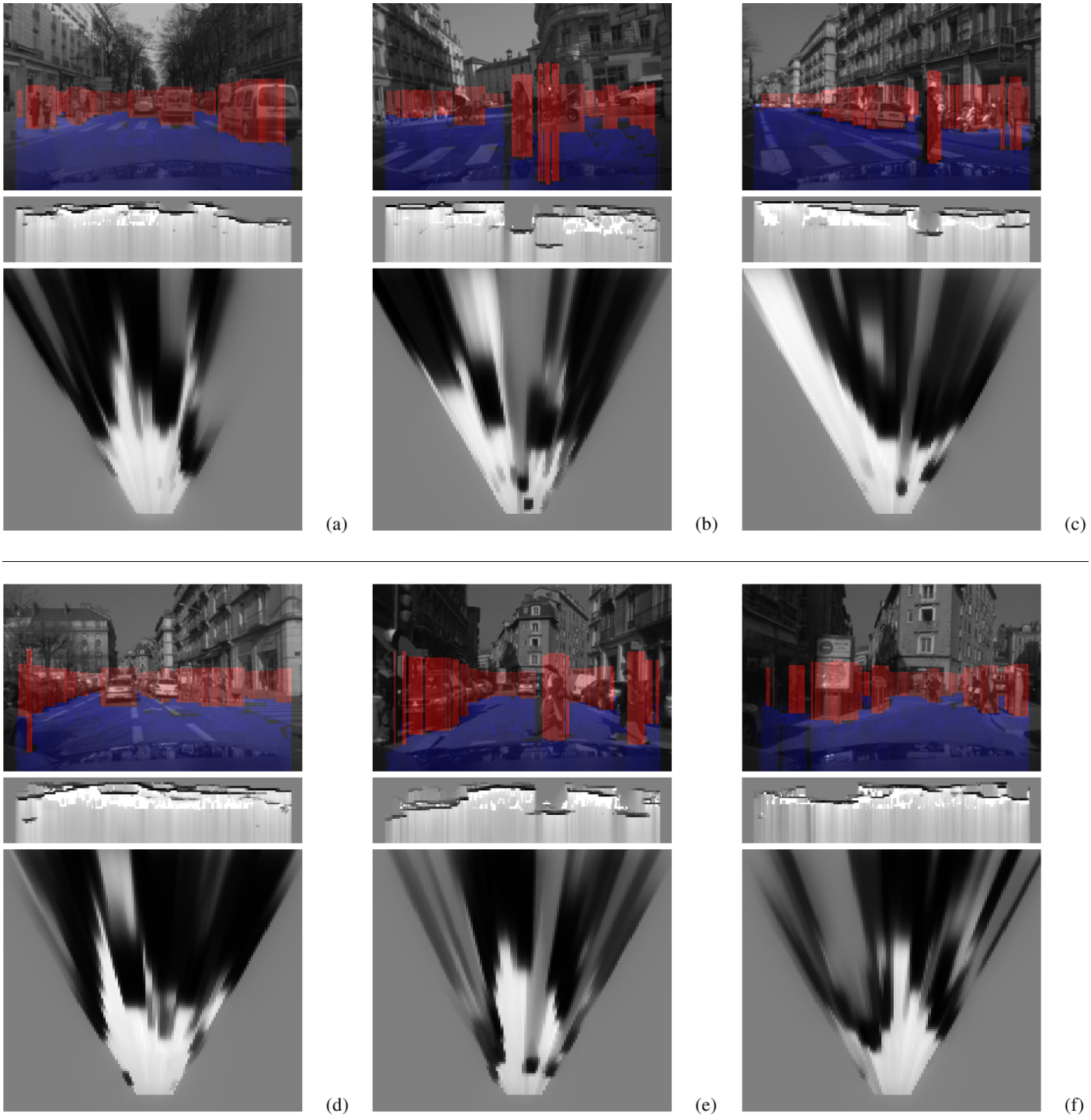


Fig. 7. Example results obtained for a sequence of stereo data. For each case: (up) the left image of the stereo pair, with a representation of the free space, obtained by thresholding the u-disparity grid, (center) the u-disparity grid, (bottom) the Cartesian grid, after filtering with the BOF.

value greater than 0.75 are represented in red (with an arbitrary height of 1.8m). The regions of the image with an occupancy value lower than 0.4 are represented in blue, projected on the detection plane. The central image is the corresponding u-disparity grid. The lower image is the associated Cartesian grid, after filtering with the BOF.

It appears that our approach provides, in most cases, a correct estimation of the free space. In the image, the decomposition can be very precise, even in complicated scenes containing pedestrians, cars, poles, walls, bicycles, etc.

You can note that the regions of the road surface, where some strong texture information is present (e.g. markings or crosswalk), improve the estimation of the grid by lowering the occupancy probability. Partially occluded objects are visible in the grid, for example the motorcycle behind the

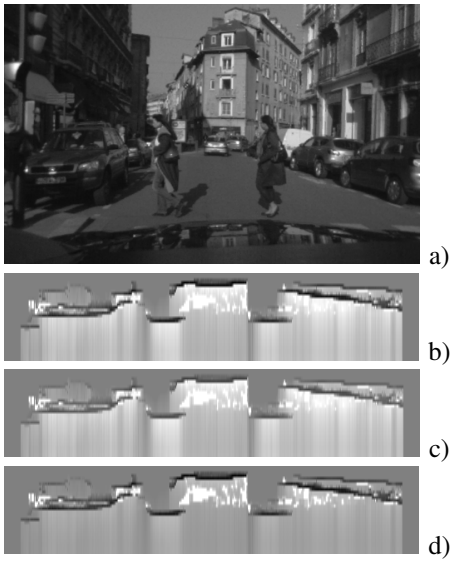


Fig. 8. *Influence of the P_{FP} and P_{FN} parameters: a) left image from the stereo pair, b) standard parameters: $P_{FP} = P_{FN} = 0.02$, c) a higher value of P_{FP} (0.25) reduces the confidence in occupied (dark) areas, d) a higher value of P_{FN}(0.25) reduces the confidence in empty (light) areas.*

pole (Fig. 7-b).
The thresholding used for the color visualization gives a good first approximation of the scene decomposition, but it is sensitive to the threshold value (blank space appears in the middle of objects). In order to retrieve an accurate object level representation, which can be useful for risk estimation, a more advanced grid clustering algorithm would be necessary.

Figure 8 shows that the value of the parameters P_{FP} and P_{FN} is not critical for the algorithm. Changing these values allows to adjust the confidence in occupied/free areas, but will not change the nature of the results. For these reason, we could safely make the assumption that these parameters are constant, and empirically set a value. On another hand, we could refine the approach by considering relationships between system parameters (e.g. the probability of a false negative P_{FN}, is related to the density of the disparity image).

Cartesian grids: Figure 7 also shows the remapped Cartesian occupancy grids, filtered with the Bayesian Occupancy Filter. The resulting description of the environment is good for short distances, but the precision becomes poor beyond 15m. This is not due to the method itself, but rather to the parameters of our sensor, which are adapted for applications at low speed, in urban environments. Cameras with longer focal length or a finer resolution would be necessary for high speed applications. For this experimental platform, fusion with longer-range lidar is done for longer-range accuracy [24].

In order to validate the results with a known reference, we computed occupancy grids with two four-layers laser scanner to compare them with the Cartesian grids obtained using stereo-vision. Figure 9 illustrates this on two typical urban cases. The wider field of view of the grids obtained using laser scanners (c,f) is explained by the combination of two lidar, whose scanning areas are only partially overlapping.
On the first example (a), one can see that the free space of the street is correctly mapped using the stereo approach (b). It is worth noticing that the classical L-shape of the white van in the lidar scan (c) is also visible in the stereo grid (b). The second example (d) illustrates that the perception of near objects is very similar with both approaches. Thus, both the pedestrians and the scooter are clearly and precisely visible in both grids (e,f), at the same position. The limitation of the resolution of our sensor is clearly visible on this case: the building in the background is mapped to large areas (e.g. upper left part of the stereo grid) (e).

C. Implementation on GPU

In order to be reactive enough for this application, all the tasks have to be done at video frame-rate (*i.e.* less than $30ms$). Our approach has been designed to be highly parallel, so that using a GPU with many execution cores is very appropriate. The NVIDIA CUDA API offers a convenient way to use such a system. Here are some implementation choices that we made to gain computational efficiency.
For the matching stage, disparity values are processed sequentially, by shifting the right image of the stereo pair. For each disparity value, the aggregation of the cost, using a correlation window, is done individually and in parallel for each pixel. The images are loaded in texture memory, to benefit from the prefetch capability of GPUs.
The u-disparity occupancy grid is computed by processing all the columns in parallel, since all the columns are independent. Then, the local maximum is computed for each group of pixels which correspond to a cell of the Cartesian grid. Thus, the subsequent remapping can be done by processing all the cells of the Cartesian grid in parallel. Finally, note that all the processing chain is done on GPU, so that no exchanges with the central memory of the computer are necessary.
The approach relies at different levels on the computation of histograms (u-disparity image, histograms of visible/ possible/ observed pixels). Therefore, atomic operations are used to speed-up the computation.

Computation time: The mean computation time for each stage of the algorithm is:

- stereo matching: 6ms
- u-disparity grid: 0.1ms
- inverse mapping: 0.1ms

Thanks to the GPU implementation, our approach can run in real time, including the stereo matching. By comparison, the same processing chain requires about 150ms on CPU, with optimizations. The orders-of-magnitude improvement on the GPU is due to the fact that our approach allows highly parallel computation. Moreover, the complexity is not related to the content of the scene, so the computation time remains constant over time.

VII. Conclusion and Future Work

This paper has presented a novel stereo-vision algorithm to create an occupancy grid, which provides four advantages.

hal-00671208, version 1 - 16 Feb 2012

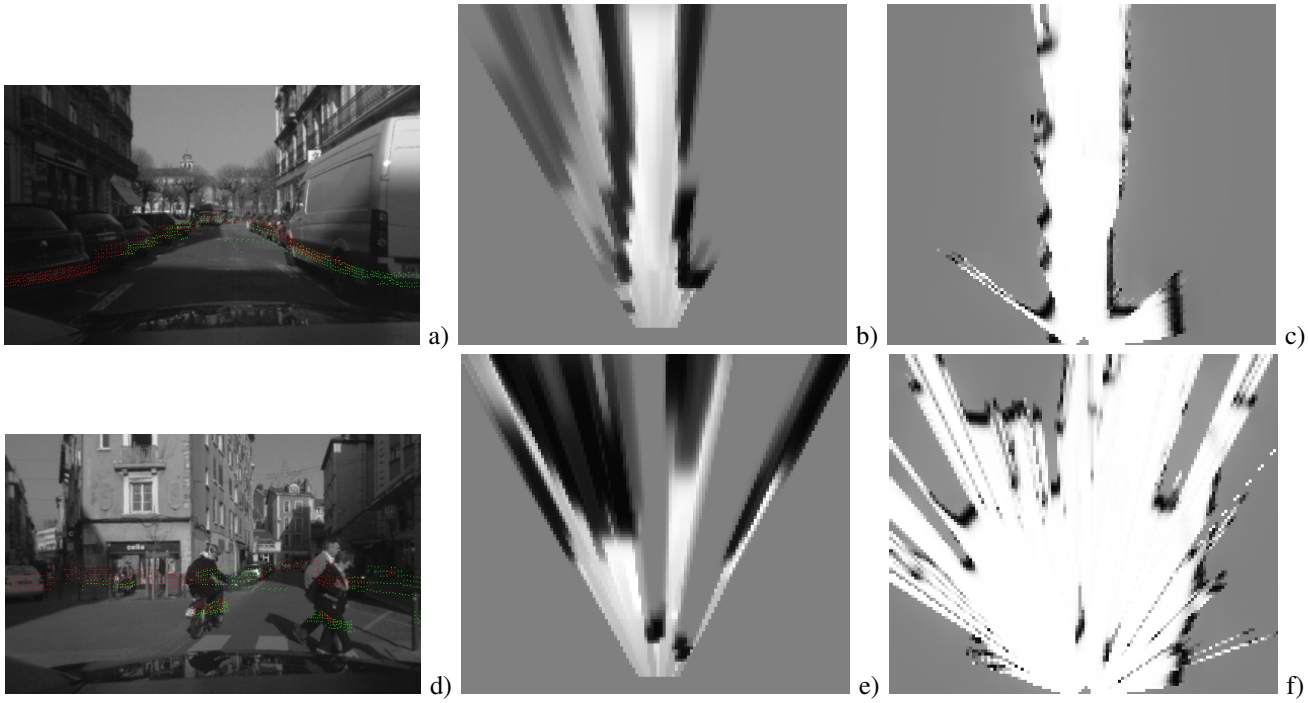


Fig. 9. *Comparison with grids obtained using two four-layers laser scanners. a& d) left image of the stereo pair, with laser impacts projected on it. Red and Green impacts correspond respectively to the left and right laser scanner. b& e) Cartesian occupancy grids computed using stereo-vision, as described in the paper. c& f) Occupancy grids computed using the laser scanners.*

First, it provides a realistic probabilistic representation of the environment, which can deal with partially occluded areas of the scene. Second, real-time operation is possible because using the disparity space allows computationally efficient and parallel calculations. Third, since visibility calculations are simpler in the u-disparity space, we presented a formal means of calculating the occupancy as a function of 3 variables:

- the probability that an obstacle is visible;
- the confidence in the observation of the obstacle;
- the confidence that the roadway was observed.

Finally, knowledge about the uncertainty of stereoscopic sensors is used to create a smooth, filtered, Cartesian occupancy grid. In summary, the main contribution of this paper is to calculate occupancy grids from stereo images in a computationally efficient way, which formally accounts for the probabilistic nature of the sensor. This approach has been tested in real road conditions, with promising results. Particularly, combined with the Bayesian Occupancy Filter, it can be used for obstacle detection, by means of clustering in the occupancy and velocity grids.

There is work planned to improve this technique. First, we plan to test this algorithm extensively, in conjunction with laser scanners for Bayesian sensor fusion, and early results are already presented in [24]. Then, adding sub-pixel estimation of the disparity values will provide improved accuracy, without modifying the method itself. The problem is to use an approach which provides actual separation between the road surface and the vertical objects.

ACKNOWLEDGMENT

This work has been done within the context of the Arosdyn project[1] on collision risk estimation [25]. The authors would like to thank Dr. Pierre Bessiere for his review of this probabilistic model, as well as Nicolas Turro and Jean-François Cuniberto for their work on the experimental platform. The authors also thank Toyota Motor Europe for their support of our experimental work on the vehicle. Dr Yoder would like to thank Inria and Ohio Northern University for allowing him to spend a year with the E-motion team at Inria Rhône-Alpes.

REFERENCES

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, 2005.
[2] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, 1989.
[3] C. Coue, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessiere, "Bayesian occupancy filtering for multi-target tracking: an automotive application," *Int. Journal on Robotics Research*, vol. 25, January 2006.
[4] M.-K. Tay, K. Mekhnacha, C. Chen, M. Yguel, and C. Laugier, "An efficient formulation of the bayesian occupation filter for target tracking in dynamic environments," *Int. Journal of Autonomous Vehicles*, 2008.
[5] T. Gindele, S. Brechtel, J. Schrder, and R. Dillmann, "Bayesian occupancy grid filter for dynamic environments using prior map knowledge," in *IEEE Intelligent Vehicles Symp.*, (Xian, China), 2009.
[6] R. Danescu, F. Oniga, and S. Nedevschi, "Modeling and tracking the driving environment with a particle-based occupancy grid," *IEEE Trans. Intelligent Transportation Systems*, vol. 12, pp. 1331 –1342, dec. 2011.
[7] R. Labayrade, D. Aubert, and J. Tarel, "Real time obstacles detection on non flat road geometry through v-disparity representation," in *IEEE Intelligent Vehicles Symp.*, (Versailles, France), 2002.
[8] C. Caraffi, S. Cattani, and P. Grisleri, "Off-road path and obstacle detection using decision networks and stereo vision," *IEEE Trans. Intelligent Transportation Systems*, vol. 8, pp. 607 –618, dec. 2007.

[1] http://arosdyn.gforge.inria.fr/

hal-00671208, version 1 - 16 Feb 2012

hal-00671208, version 1 - 16 Feb 2012

[9] S. Nedevschi, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, T. Graf, and R. Schmidt, "High accuracy stereovision approach for obstacle detection on non planar roads.," in *IEEE Int. Conf. on Intelligent Engineering Systems*, (Cluj Napoca, Romania), 2004.

[10] U. Franke and A. Joos, "Real-time stereo vision for urban traffic scene understanding," in *IEEE Intelligent Vehicles Symp.*, (Dearborn, USA), 2000.

[11] K. Huh, J. Park, J. Hwang, and D. Hong, "A stereo vision-based obstacle detection system in vehicles," *Optics and Lasers in Engineering*, vol. 46-2, 2008.

[12] V. Lemonde and M. Devy, "Obstacle detection with stereovision.," in *Mechatronics and robotics*, (Aachen, Germany), 2004.

[13] T. Veit, "Connexity based fronto-parallel plane detection for stereovision obstacle segmentation," in *IEEE Int. Conf. on Robotics and Automation, workshop*, (Kobe, Japan), 2009.

[14] C. Braillon, C. Pradalier, K. Usher, J. Crowley, and C. Laugier, "Occupancy grids from stereo and optical flow data," in *Int. Symp. on Experimental Robotics*, (Rio de Janeiro, Brazil), 2006.

[15] L. Matthies and A. Elfes, "Integration of sonar and stereo range data using a grid-based representation," in *IEEE Int. Conf. on Robotics and Automation*, (Philadelphia, USA), 1988.

[16] D. Murray and J. Little, "Using real-time stereo vision for mobile robot navigation," *Autonomous Robots,*, vol. 8, January 2000.

[17] H. Badino, U. Franke, and R. Mester, "Free space computation using stochastic occupancy grids and dynamic programming," in *IEEE Int. Conf. on Computer vision, Workshop on Dynamical Vision*, (Rio de Janeiro, Brazil), 2007.

[18] D. Pfeiffer and U. Franke, "Modeling dynamic 3d environments by means of the stixel world," *IEEE Intelligent Transportation Systems Magazine*, vol. 3, pp. 24 –36, fall 2011.

[19] M. Perrollaz, A. Spalanzani, and D. Aubert, "A probabilistic representation of the uncertainty of stereo-vision and its application to obstacle detection," in *IEEE Intelligent Vehicles Symp.*, (San Diego, USA), 2010.

[20] Z. Hu, F. Lamosa, and K. Uchimura, "A complete u-v-disparity study for stereovision based 3d driving environment analysis," in *IEEE 3DIM*, (Washington, USA), 2005.

[21] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision (second edition)*. Cambridge University Press, 2003.

[22] P. Burt, L. Wixson, and G. Salgian, "Electronically directed "focal" stereo," in *IEEE Int. Conf. on Computer Vision*, (Los Alamitos, USA), 1995.

[23] M. Bertozzi, A. Broggi, A. Fascioli, and R. Fascioli, "Stereo inverse perspective mapping: Theory and applications," *Image and Vision Computing Journal*, vol. 8, 1998.

[24] J.-D. Yoder, M. Perrollaz, I.-E. Paromtchik, Y. Mao, and C. Laugier, "Experiments in sensor fusion using the bayesian occupancy filter," in *Int. Symp. on Experimental Robotics*, (Delhi, India), 2010.

[25] C. Laugier, I. Paromtchik, M. Perrollaz, Y. Mao, J.-D. Yoder, C. Tay, K. Mekhnacha, and A. Nègre, "Probabilistic Analysis of Dynamic Scenes and Collision Risk Assessment to Improve Driving Safety," *Intelligent Transportation Systems Magazine*, vol. 3, no. 4, 2011.



Mathias Perrollaz graduated from National Polytechnic Institute of Grenoble (INPG) in 2003, with a major in signal and image processing. He started working on ITS at the Images and Signals Laboratory in Grenoble (CNRS), and then attended the perception team of the LIVIC (INRETS). He received the Ph.D. degree from Paris-6 university (UPMC) in 2008, for his work on multi-sensor obstacle detection. Since April 2009, he is working at Inria Grenoble (France) on probabilistic methods for ITS. Between may and september 2011, he worked on perception for robotic manipulators at Ohio Northern University, USA. Mathias Perrollaz also taught in Paris-10, Grenoble-1, INPG and Grenoble-2 Universities.
His main research interests are computer-vision, ITS, and perception for robotics.



John-David Yoder graduated from the University of Notre Dame, Notre Dame, IN, USA with the B.S. in mechanical engineering in 1991. He received his M.S. and Ph.D. in mechanical engineering also from Notre Dame in 1996 and 1999 respectively.

He is Associate Professor and Chair of the mechanical engineering department at Ohio Northern University, Ada, OH. He has previously served as Proposal Engineer and Proposal Engineering Supervisor at Grob System, Inc. and Software Engineer at Shaum Manufacturing, Inc. He has held a number of leadership and advisory positions in various entrepreneurial ventures. He is currently a KEEN (Kern Entrepreneurial Education Network) Fellow, and for the 2009-10 academic year is an invited professor at Inria Rhône-Alpes, Monbonnot, France.
Research interests include computer vision, mobile robotics, intelligent vehicles, entrepreneurship, and education.



Amaury Nègre is Research Engineer at CNRS since 2008. He graduated from the National Polytechnic Institute of Grenoble (INPG), France in 2005, with a major in image processing. Then he received his PhD degree in vision and robotics from INPG in 2009 for his work on vision based obstacle detection for mobile robots. Now he is with the Grenoble Informatics Laboratory (LIG) and offers engineering support in experimental robotics and ambient intelligence.



Anne Spalanzani received her PhD in Cognitive and Computer Science from the University Joseph Fourier of Grenoble, France, in 1999. She is associate professor in e-motion team of Inria and the University Pierre Mendès France of Grenoble since 2003. She works on perception and autonomous navigation integrating probabilistic prediction of risk of collision and social interactions.



Christian Laugier is Research Director at Inria and Scientific Leader of the e-Motion Team. He is also responsible at INRIA for scientific relations with Asia & Oceania. He received the PhD degree in computer science from Grenoble University, France in 1976. His current research interests mainly lie in the areas of motion autonomy, intelligent vehicles and probabilistic robotics. He has co-edited several books in the field of robotics and several special issues of scientific journals such as IJRR, Advanced Robotics, JFR, or IEEE Transactions on ITS. In 1997, he was awarded the Nakamura Prize for his contributions to "Intelligent Robots and Systems". Dr. Laugier is a member of several scientific committees including the Steering/Advisory Committees of the IEEE/RSJ IROS, FSR, and ICARCV conferences. He is also co-Chair of the IEEE RAS Technical Committee on AGV & ITS. He has been General Chair or Program Chair of such conferences as: IEEE/RSJ IROS'97, IROS'02, IROS'08, or FSR'07. Additionally to his research and teaching activities, he co-founded four start-up companies in the fields of robotics, computer vision, computer graphics, and Bayesian programming tools. He has served as scientific consultant for the ITMI, Aleph Technologies, and ProBayes companies.