HAL
open science

Using the disparity space to compute occupancy grids from stereo-vision

Mathias Perrollaz, John-David Yoder, Anne Spalanzani, Christian Laugier

To cite this version:

Mathias Perrollaz, John-David Yoder, Anne Spalanzani, Christian Laugier. Using the disparity space to compute occupancy grids from stereo-vision. International Conference on Intelligent Robots and Systems (IROS), Oct 2010, Taipei, Taiwan. inria-00527785

HAL Id: inria-00527785

https://inria.hal.science/inria-00527785

Submitted on 20 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using the disparity space to compute occupancy grids from stereo-vision

Mathias Perrollaz, John-David Yoder, Anne Spalanzani and Christian Laugier

***Abstract*— The occupancy grid is a popular tool for probabilistic robotics, used for a variety of applications. Such grids are typically based on data from range sensors (e.g. laser, ultrasound), and the computation process is well known [1]. The use of stereo-vision in this framework is less common, and typically treats the stereo sensor as a distance sensor, or fails to account for the uncertainties specific to vision. In this paper, we propose a novel approach to compute occupancy grids from stereo-vision, for the purpose of intelligent vehicles. Occupancy is initially computed directly in the stereoscopic sensor's disparity space, using the sensor's pixel-wise precision during the computation process and allowing the handling of occlusions in the observed area. It is also computationally efficient, since it uses the u-disparity approach to avoid processing a large point cloud. In a second stage, this disparity-space occupancy is transformed into a Cartesian space occupancy grid to be used by subsequent applications. In this paper, we present the method and show results obtained with real road data, comparing this approach with others.**

I. Introduction

Creating (or having *a priori*) a model of the local environment has always been one of the requirements for successful implementation of a mobile robot. One common approach has been to build an occupancy grid. These grids are often created based on data from one or more range sensors, such as laser or ultrasound. Because of the imperfection in such sensors, it is current practice to create the grid by using a probabilistic approach [1][2]. While stereo camera pairs have been widely used on mobile robots, their use for the creation of occupancy grids has been somewhat less common. In some cases, researchers have used the stereo cameras purely as a distance sensor, and used the same approach as with a laser sensor [3]. Others have used stereo-specific methods, but have not completely considered the nature of the stereoscopic data. In [4] or [5], the authors only consider the first detected object for each column and suppose that it occludes the field ahead.

This paper presents a novel approach for the construction of occupancy grids using a stereo camera pair, 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 idea is not to map the environnement at a global scale, but rather to estimate the free space in front of the vehicle. Also, processing time is critical for this application, as the system will be required to execute in real-time. Finally, it is important to see all obstacles, even those which may be behind another obstacle, such as a pedestrian stepping from behind a car. For example, if a car is parked on the curb, it may present no risk to the ego-vehicle. But a person stepping out from behind the car would present a risk - and by the time the person is no longer occluded behind the car, it may be too late to safely stop the vehicle. One of the critical advantages of cameras over 2-D laser sensors for this application is precisely the ability to see such obstacles. A 3D laser sensor has similar advantages, but is much more expensive and slower than a vision sensor.

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 also formally considers the visibility of the different regions of the image in the calculation, and thus can deal with partially-occluded objects. Because of its use of the u-disparity space, it is computationally efficient. Finally, the method formally considers the reduction in accuracy of the disparity image with distance from the sensor. The paper will detail the methodology, and show results with data from real images obtained on the road.

Section II provides a brief review of the use of the disparity space, with specifics related to the intelligent vehicle application. Section III explains the overall methodology. Section IV shows results, and compares the method with other approaches. Finally, Section V discusses future work and conclusions.

II. 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. It should be noted that this application only calls for the visual sensor to create an instantaneous occupancy grid of the area in front of the vehicle.

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.

M. Perrollaz (mathias.perrollaz@inrialpes.fr), A. Spalanzani (anne.spalanzani@inrialpes.fr) and C. Laugier (christian.laugier@inrialpes.fr) are with the INRIA Grenoble.

J-D. Yoder (j-yoder@onu.edu) is with Ohio Northern University, currently invited professor at INRIA Grenoble.

For simplicity in notation, and without loss of generality, 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 [6]:

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

B. The u-disparity approach

1) The idea: The u-disparity approach [7] is a complement to the v-disparity originally described in [8]. 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, as illustrated on figure 1. On this image, the lower lines correspond to close areas, while the upper lines contain information about the long distances. Vertical objects (*i.e.* with constant disparity, *e.g.* the pedestrians) appear as portion of straight lines.

Fig. 1. *Left image from a stereo pair (top), the associated disparity image (mid) and u-disparity representation of the scene (bottom).*

2) The detection plane: For occupancy grid computation, we have to consider a detection plane $\mathcal{P}_D$, that is the support for the grid. As shown on figure 2, $\mathcal{P}_D$ is chosen to be parallel to the plane defined by the baseline and the optical axes. This constaint is possible, since the objective is to build an occupancy grid in the vehicle's frame.
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 $z_w = 0$.

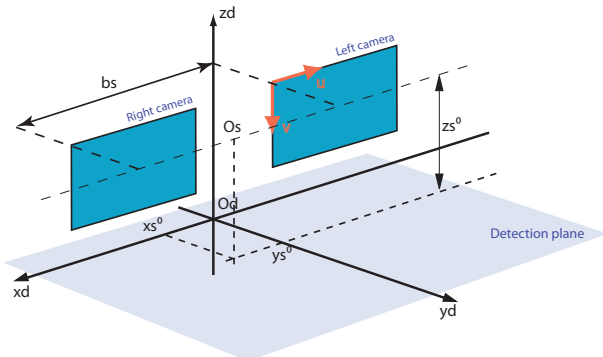

Fig. 2. *Simplified geometrical configuration of the stereoscopic sensor, and reference to the common coordinate system R_D.*

Considering equation 1, it appears 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 remainder 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 observation function G_D:

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

with:

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

3) The alignment of rays: Another advantage of the u-disparity approach for occupancy grid computation is that, contrary to the cartesian representation, it lets appear as parrallel the rays of light that go through the camera matrix. Indeed, a set of vertically aligned rays is represented by a column in the u-disparity image. In a similar way, the lines of the v-disparity image correspond to sets of horizontally aligned rays. This behavior is promising since it allows easy estimation of occluded areas.

C. Road-obstacle separation

For the reminder of the paper, it is assumed that pixels can be distinguished as being from the road surface or from the obstacles. There are several methods to do this, such as estimating the road surface and thresholding the height of the pixels, as in [3]. We choose to use a double correlation framework, which exploits different matching hypotheses for vertical and horizontal objects, as described in [9] and detailed in [6]. It provides immediate classification of the pixels during the matching process. 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.

III. Occupancy grid computation from 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, this allows direct consideration of visible and occluded portions of the image. Second, this approach allows 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. Estimation of the occupancy of a cell

We seek to calculate $P(O_U)$, the probability that a cell U is occupied by an obstacle. This probability will depend on the visibility, V_U, and on the confidence on the observation of an obstacle, C_U. V_U and C_U are binary random variables (*i.e.* $P(V_U = 1)$ measures the probability that the cell U is visible, $P(C_U = 1)$ is the probability that an obstacle has been observed in cell U). Experience provides knowledge about 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. Indeed, some boundary conditions of $P(O_U|V_U, C_U)$ are known. For example, if the cell is not visible, nothing is known about its occupancy, so:

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

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

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

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

$$P(O_U|V_U = 1, C_U = 0) = P_{FN} \tag{6}$$

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 arbitrarily set according to the sensor's features. Finally, the laws of probability are used to obtain the full decomposition of $P(O_U)$:

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

V_U and C_U being boolean variables, this means:

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

So to compute the occupancy probability, it is necessary to estimate the values $P(V_U = 1)$ and $P(C_U = 1)$ with respect to the disparity data.

Let us define the maximum detection height h. For a given cell U of the grid, we define the number of possible measurement points as:

$$N_P(U) = v_0(d) - v_h(d) \tag{9}$$

$v_0(d)$ 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. Considering equation 1:

$$N_P(U) = \frac{\alpha_v}{\alpha_u} \cdot \frac{h.d}{b_s} \tag{10}$$

The number of observed obstacle pixels for the cell is the number of possible pixels whose disparity value is d. It can be directly measured in the obstacle u-disparity image:

$$N_O(U) = I_U^{obst}(u, d) \tag{11}$$

1) Estimation of the visibility of a cell: To estimate the probability that a cell U is visible, we classify the pixels of the obstacle disparity image whose coordinates are $(u, v \in [v_h, v_0])$ (*i.e.* possible pixels):

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

The number of visible pixels for any cell U is $N_V(U)$. The probability of visibility is estimated to be the ratio between visible pixels and possible pixels:

$$P(V_U = 1) = \frac{N_V(U)}{N_P(U)} \tag{12}$$

2) Estimation of the confidence of observation: We choose to express the confidence on the observation of an obstacle as a function of the ratio:

$$r_O(U) = \frac{N_O(U)}{N_V(U)} \tag{13}$$

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:

$$P(C_U = 1) = 1 - e^{-\frac{r_O(U)}{\tau_O}} \tag{14}$$

where τ_O is a constant. Note that although C_U is a binary variable (either true or false), this probability density function is continuous.

3) Resulting probability density function: Figure 3 illustrates equation 8, the probability density function of occupancy with respect to the visibility and to r_O. For this plot, parameters are set to: $P_{FP} = 0.01$, $P_{FN} = 0.05$ and $\tau_O = 0.15$. These parameters were set manually.

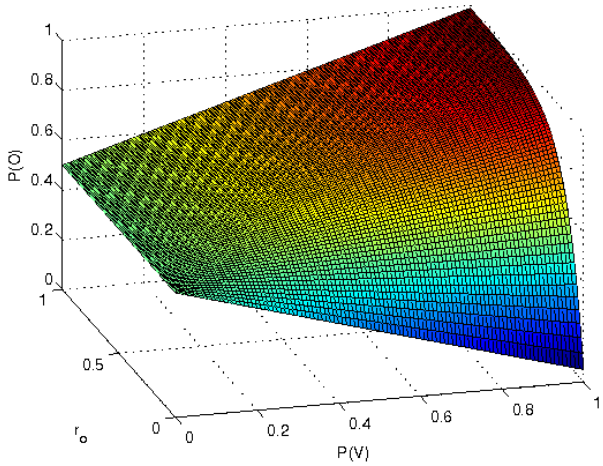

Fig. 3. *Probability density function* $P(O_U)$.

C. Resulting grid in the u-disparity plane

Knowing the probability density function $P(O_U|V_U, C_U)$, the computation of the occupancy grid in the u-disparity image is straightforward. Figure 4 shows a detail of a scene and the associated occupancy grid in the u-disparity plane. Regions with obstacles appear in light color, while free space appears in dark color. It is noticeable that regions situated in front of obstacles are seen as free. Thanks to our approach which formally considers the visibility, the pedestrian is detected even if it is partially occluded by the box.

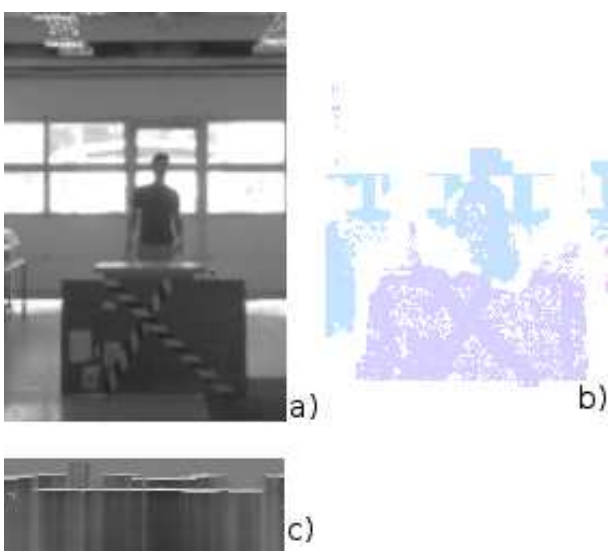

Fig. 4. *Detail of a scene (a), associated disparity map (b) and occupancy grid in u-disparity. The box and the pedestrian behind the box are both apparent on the u-disparity occupancy grid. This would not have been possible without considering the visibility.*

D. Computation of the occupancy grid in Cartesian space

The cartesian occupancy grid 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 $\mathrm{S}_U(U)$ of a pixel $U = (u, d)$ 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: $\mathrm{S}_X(U) = G(\mathrm{S}_U(U))$. To compute the occupancy grid, the occupancy probability of a pixel U is simply attributed to the area $\mathrm{S}_X(U)$ of the detection plane. For short distances, several pixels can have an influence on the same cell of the metric grid. Therefore, 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:

$$P(O_X) = max(P(O_U)/X \in \mathrm{S}_X(U)) \tag{15}$$

IV. Experimental results

A. Experimental setup

The algorithm has been evaluated on sequences from the French LOVe project [10], dealing with pedestrian detection. The images are taken with a pair of SMAl CMOS cameras, which provide images every 30 ms. They are reduced to quarter VGA resolution (320*240 pixels) before the matching process. The length of the stereo baseline is 43 cm.
The observed region is $-7.5m < x_d < 7.5m$ and $0m < y_d < 35m$, maximum height is $h = 2.0m$, and the cell size is $0.25m*0.25m$. The correlation window measures $7pixels$ in width and $19pixels$ in height.
The occupancy grid computation parameters are set to: $P_{FP} = 0.01$, $P_{FN} = 0.05$ and $\tau_O = 0.15$.

B. Results

Figure 5 shows typical results. For each of the cases (a-f), the left camera's image is shown in the top left, the u-disparity occupancy grid in the bottom left, and the cartesian occupancy grid on the right. For the occupancy grids, lighter colors indicate an increased likelihood of occupation, while the background color represents a probability of occupation of 0.5 (which means that there is no knowledge about the likelihood of occupation of that cell).

In case a), the closest obstacle is the truck to the right. In both occupancy grids, we can see the truck to the right, and a variety of obstacles at a further distance in front and to the left. In case b), we see a more cluttered scene, and in the cartesian grid, the various obstacles are visible. Left to right in the grid, note the oncoming car, the pedestrian further away, the stoplight pole in the center, the truck as a large obstacle further away, and the car on the far right. Case c) shows that obstacles are seen at longer distance. Note that the area between the camera and the other vehicles is seen as free space. Case d) shows one of the advantages of this method, as we can identify both the car turning in front (the bottom right of the cartesian grid), and the sign behind it. This effect is easily seen in the u-disparity grid, on the right, where one can clearly see two obstacles. The algorithm also correctly notes that there is some likelihood that the area between the car and the sign is unoccupied. Cases e) and f) show two scenes cluttered with pedestrian traffic. In the u-disparity grid, it is obvious that objects are

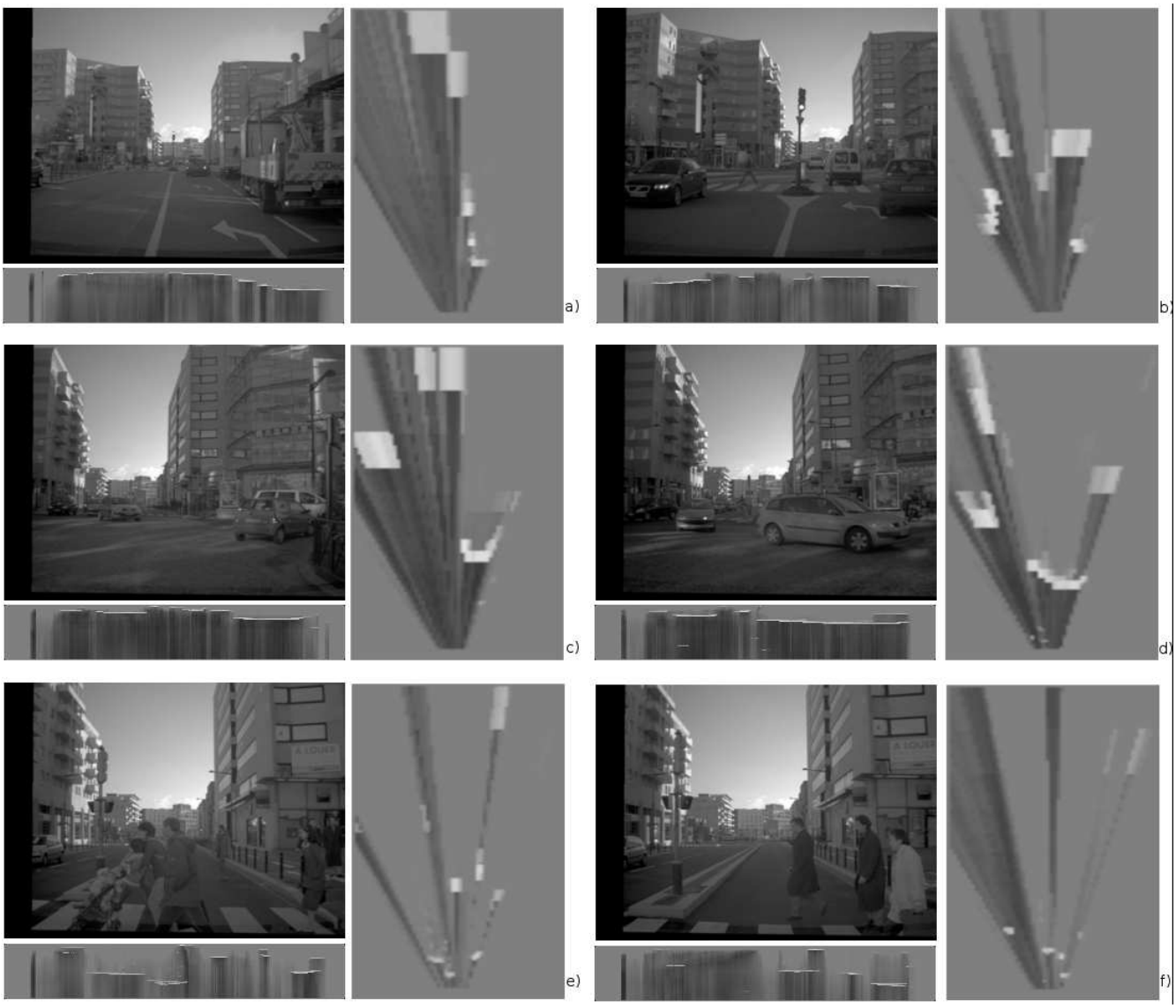

Fig. 5. *Typical results obtained with a dataset from the French project LOVe. For each result: the left image of the stereo pair, the occupancy grid in the u-disparity plane and the occupancy grid in cartesian space.*

detected at a wide variety of distances. In the cartesian grid, note the pedestrians, as well as a few of the vertical posts on the right. Because the obstacles are close to the vehicle, most of the occupancy grid in these cases remains unknown.

Figure 6 compares this approach to two other methods of using a stereoscopic sensor to create an occupancy grid, using case c from figure 5. Method a) detects the maximum disparity of each column (as in [5]) occupancy probability is P_{FN} for higher disparity values, and 0.5 behind the detected object. This approach is very sensitive to noise in the disparity map and can not percieve partially occluded parts of the scene. Method b) relies only on filling the grid based on knowledge of the road and obstacles pixels, as described in [6]. As such, much of the image has no information (leading to large areas of 0.5 occupancy probability), and the resulting grid tends to have high certainty. Finally, the method described in this paper is shown in c). The method formally takes into account the probability of false measurements, making it less susceptible to noise, and finds the obsacle behind the car whereas method a) found only the part of the obstacle in front of the car. The probabilistic model has also provided a more realistic variation in the occupancy probablity values.

C. Computation time

On a standard laptop, without optimizations, the algorithm for the u-disparity occupancy grid computation runs in 10 ms. The conversion into cartesian grid runs in 40 ms. Considering that all the columns of the occupancy grid in u-disparity can be processed independantly, many operations can be performed in parallel. Thus it can be efficiently implemented on a parellel architecture, *e.g.* a GPU, to furter reduce computation time. We are now working on such implementation of the algorithm on GPU, using NVIDIA's

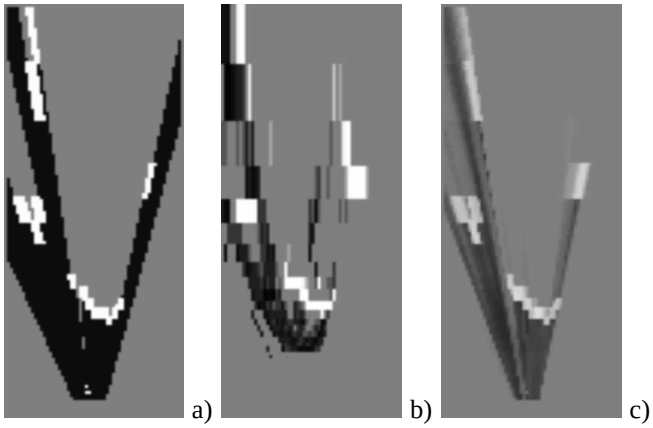

Fig. 6. *Comparison of three approaches to compute the occupancy grid. a) detection of the maximum disparity of each colum, b) exploitation of obstacle and road u-disparity images as measurement point, c) computation in the u-disparity plane.*

CUDA. Preliminary tests showed that it is possible to run the complete algorithm at video framerate (25fps), from the stereo matching to the construction of the cartesian grid.

V. Conclusion and future work

The algorithm described in this paper offers several advantages over other approaches using stereo vision to create an occupancy grid. First, using the u-disparity as a starting point allows computationally efficient calculations. Second, computing the initial occupancy grid in the u-disparity space allows for simpler calculation of the visibility. It also allows the use of data with a pixel-wise precision all along the computation process. Third, we have presented a formal means of considering the occupation as a function of the probability that an obstacle is visible, and of the confidence in the observation of the obstacle. In summary, the 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.

Much work remains to be done. It is expected that this algorithm will soon be tested on a vehicle also equipped with dual laser range finders. The three sensors (2 lasers and stereo-vision) will all provide occupancy grids to a Bayesian Occupancy Filter. Comparing and fusing these occupancy grids will provide another means of evaluating the output of this algorithm. Third, the method will be updated to formally consider the semi-occlusions in the stereoscopic images (*i.e.* pixels next to disparity discontinuities). We also hope to refine the approach by considering relationships between parameters in the system (for example, the probability of a false negative, or false positive, is related to the density of the disparity image). Finally, work is being done to filter the Cartesian grid based on the uncertainty of the stereo sensor, and to take advantage of pixels that are identified as road pixels.

Acknowledgment

This work has been done within the context of the Arosdyn project. The authors would like to thank Dr. Pierre Bessiere for his review of this probabilistic model and Dr. Amaury Negre for its work on the GPU implementation. Dr Yoder would like to thank INRIA and Ohio Northern University for allowing him to spend this year with the E-motion team at INRIA Grenoble.

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,” *IEEE Computer*, vol. 22-6, 1989.
[3] C. Braillon, C. Pradalier, K. Usher, J. Crowley, and C. Laugier, “Occupancy grids from stereo and optical flow data,” in *International Symposium on Experimental Robotics (ISER)*.
[4] L. Matthies and A. Elfes, “Integration of sonar and stereo range data using a grid-based representation,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 1988.
[5] D. Murray and J. Little, “Using real-time stereo vision for mobile robot navigation,” *Autonomous Robots,*, vol. 8, January 2000.
[6] 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 Symposium (IV)*, 2010.
[7] Z. Hu, F. Lamosa, and K. Uchimura, “A complete u-v-disparity study for stereovision based 3d driving environment analysis,” in *International Conference on 3-D Digital Imaging and Modeling(3DIM)*, 2005.
[8] R. Labayrade, D. Aubert, and J. Tarel, “Real time obstacles detection on non flat road geometry through v-disparity representation,” in *IEEE Intelligent Vehicles Symposium (IV)*, 2002.
[9] P. Burt, L. Wixson, and G. Salgian, “Electronically directed ”focal” stereo,” in *IEEE International Conference on Computer Vision (ICCV)*, vol. 0, 1995.
[10] “Love : http://love.univ-bpclermont.fr/.”
[11] C. Coue, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessiere, “Bayesian occupancy filtering for multi-target tracking: an automotive application,” *International Journal of Robotics Research (IJRR)*, vol. 25, January 2006.