Using obstacles and road pixels in the disparity-space computation of stereo-vision based occupancy grids

Mathias Perrollaz, John-David Yoder and Christian Laugier

Abstract— **Occupancy grids have been used for a variety of applications in the field of robotics. These 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 less common. This paper will detail a novel approach to compute occupancy grids, as applied to intelligent vehicles. Occupancy is initially computed directly in the stereoscopic sensor's disparity space, 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. The occupancy calculation formally accounts for the detection of obstacles and the road 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 a filtering step to reduce discretization effects and explicitly account for the relation between range and uncertainty in stereoscopic data. In this paper, we present the method and show the results obtained with real road data.**

I. Introduction

A model of the local environment is a basic requirement of any intelligent vehicle, and of mobile robots in general. Occupancy grids based on sensor data have long been used for this purpose, typically constructed from beam-type range sensors, such as laser or ultrasound. Because of the imperfection in such sensors, it is current practice to create this grid by using a probabilistic approach [1] [2]. The goal of the work in this paper is to create vision-based, computationally efficient occupancy grids. These occupancy grids will further be fused with laser-based occupancy grids using the Bayesian Occupancy Filter [3]. As such, we specifically want to take advantage of some of the capabilities of a visual sensor, notably the ability to see partially occluded objects -unlike a 2D laser scannner- and the ability to detect the road.

Cameras have been used for many purposes on mobile robots. The creation of occupancy grids based on stereo camera data, however, has received relatively little attention. Some authors have chosen to use cameras as distance-sensor, using exactly the same approach as with a laser [4]. Others have used stereo-specific methods, but only consider the first detected object in each column, losing the ability to detect partially occluded objects [5] [6].

This paper presents a novel approach for the construction of occupancy grids using a stereo camera pair. The method

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. Since the system must be able to operate in real-time, processing speed is critical. 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.

The method presented here provides a formal probabilistic model to calculate the probability of occupation based on the disparity space of the stereoscopic sensor. As described in [7], the method considers the visibility of different regions of the image to deal with partially occluded objects, and is computationally efficient. The work here further accounts for road pixels and obstacle pixels both being visible. Finally the method formally considers the decreasing accuracy of the stereo sensor with distance, and applies a filtering algorithm to reduce discretization effects.

Section II provides a brief review of the use of the u and v 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. Section V details the approach taken to formally consider the variation of the sensor accuracy with distance. Finally, Section VI discusses future work and conclusions.

II. The data in the disparity space

A. Geometrical considerations

The stereoscopic sensor is considered as perfectly rectified. Cameras are classically represented by a pinhole model, $(\alpha_u, \alpha_v, u_0, v_0)$ being the intrinsic parameters. The length of the stereo baseline is b_s. In the world coordinate system, denoted R_w, the center of the stereo baseline is situated at position (x^o, y^o, z^o).

A point P, of coordinate X_w, is projected onto the left and right image planes respectively on positions (u_l, v) and (u_r, v). Consequently, in the disparity space, 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 coordinates in R_w can be retrieved from the coordinates in R_Δ.

B. The u-disparity approach

1) The idea: The u-disparity approach [8] is a complement to the v-disparity originally described in [9]. 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 $\mathcal{P}_D$, that is the support for the grid. As shown on Figure 1, $\mathcal{P}_D$ 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 point P, $x_d = x_w$ and $y_d = y_w$. Arbitrarily, one can decide to set the detection plane to $z_w = 0$.
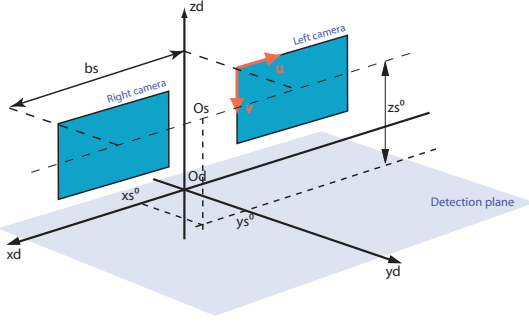

Fig. 1. *Simplified geometrical configuration of the stereoscopic sensor, and reference to the common coordinate system R_D.*

Considering such a configuration, 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 observation function $G_D : U_D \mapsto X_D$, with:

$$\begin{cases} x &= x^o + \frac{b_s}{2} + \frac{b(u-u_o)}{d} \\ y &= y^o + \frac{\alpha_u b_s}{d} \end{cases} \quad (1)$$

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. Thus, even if they intercept at the focal point of the camera, these sets of pixels are parallel to each other in the u-disparity plane. 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.

C. Road-obstacle separation

Critical to the performance of the algorithm described here is the assumption that pixels appearing in the disparity image can be distinguished as being from the road surface or from obstacles. This has been accomplished in several ways, such as estimating the road surface and thresholding the height of the pixels [4]. We choose to use a double correlation framework, which exploits different matching hypotheses for vertical and horizontal objects, as described in [10] and detailed in [11]. 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.

The basis of the method has been presented in [7]. Here we propose to improve the approach by taking into account the road pixels from the disparity image, in order to refine the estimation of the free space. The occupancy grid computation will be briefly described here for clarity. Please refer to [7] for more details.

B. Estimation of the occupancy of a cell

We seek to calculate $P(O_U)$, the probability that a cell U_D is occupied by an obstacle. This probability will depend on the visibility, V_U, and on the confidence of observation, C_U. V_U and C_U are binary random variables, and experience provides some boundaries conditions on the probability density function $P(O_U|V_U, C_U)$. If a cell is visible:

$$P(O_U|V_U = 0, C_U = c) = 0.5 \quad (2)$$

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

$$P(O_U|V_U = 1, C_U = 1) = 1 - P_{FP} \quad (3)$$

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

$$P(O_U|V_U = 1, C_U = 0) = P_{FN} \quad (4)$$

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 now assumed to be constant and known. Future work will deal with the online estimation of these parameters, considering for example the density of the disparity map.

Finally, the laws of probability are used to obtain the full decomposition of $P(O_U)$:

$$\begin{aligned} P(O_U) = \quad &P(V_U = 1) \cdot P(C_U = 1) \cdot (1 - P_{FP}) \\ &+ P(V_U = 1) \cdot (1 - P(C_U = 1)) \cdot P_{FN} \\ &+ (1 - P(V_U = 1)) \cdot 0.5 \end{aligned} \quad (5)$$

1) Estimation of the visibility of a cell: For a given cell $U_D(u, v, d)$ of the grid, let us define the number of possible measurement points as $N_P(U_D) = v_0(d) - v_h(d)$, $v_0(d)$ 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. In

order to estimate the probability that the cell U_D 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). Let d' be the estimated disparity value $d' = I_d^{obst}(u, v)$:

- if $d' > d$, the point (u, d', v) is occluded,
- if $d' = 0$, there is no observation for the ray (u, v) (*i.e.* it is not visible),
- else the pixel is said to be visible, in particular if $d' = d$ it is an observed pixel for cell U_D.

The number of visible pixels for any cell U_D is $N_V(U_D)$, the number of observed pixels is $N_O(U_D)$. We define the probability of visibility as:

$$P(V_U = 1) = N_V(U_D)/N_P(U_D) \tag{6}$$

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

$$r_O(U_D) = N_O(U_D)/N_V(U_D) \tag{7}$$

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:

$$P(C_U = 1) = 1 - e^{-\frac{r_O(U_D)}{\tau_O}} \tag{8}$$

where τ_O is a constant parameter.

C. Using the 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 R_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:

$$P(T_U) = P(O_U) \cdot P(\neg R_U) = P(O_U) \cdot (1 - P(R_U)) \tag{9}$$

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 acculumate 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 neighbourhood of U_D. Let us call $r_R(U_D)$ the ratio of non-zero road pixels in the $3 * 3$ neighbourhood of U_D. 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:

$$P(R_U) = e^{-\frac{1 - r_R(U_D)}{\tau_R}} \cdot e^{-\frac{r_O(U_D)}{\tau_O}} \tag{10}$$
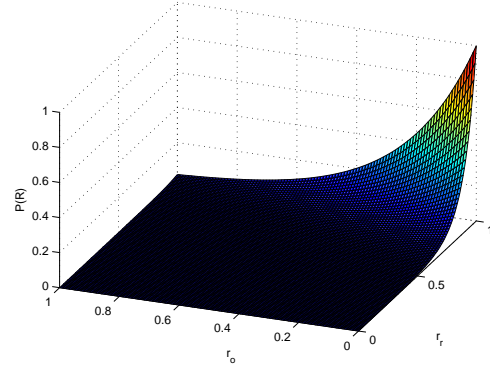

Fig. 2. *Probability density function $P(R_U)$.*

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

D. 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}, and d) shows the u-disparity occupancy grid. In b), what you see are the fronts of obstacles, resulting in (mostly) straight white 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 the back of the car, the lower image to the left is the post in the road, etc. 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 d) white pixels represent high probability of occupancy, and black very low probability. This occupancy grid maintains strong information from the obstacles (they remain white lines), while a cell is empty (black) in areas
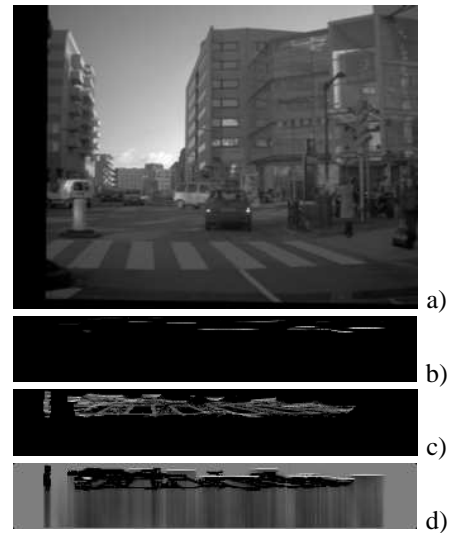

Fig. 3. *Computation of the occupancy grid in the u-disparity plane. a) left image from the stereo paire; b) obstacle u-disparity image; c) road u-disparity image; d) occupancy gird in the u-disparity plane.*

Fig. 4. *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.*

where the road was detected. Most areas behind obstacles are completely unknown, meaning they are assigned a value of $P(T_U) = 0.5$.

E. *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 $S_U(U_D)$ 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: $S_X(U_D) = G(S_U(U_D))$. To compute the occupancy grid, the occupancy probability of a pixel U_D is simply attributed to the area $S_X(U_D)$ of the detection plane.

This could have been done through inverse mappping of the pixels, but for short distances several pixels have an influence on the same cell of the Cartesian grid. Therefore, it is necessary to estimate the occupancy according to this set of data. We choose to use a max estimator, which ensures a conservative estimation of the probability of occupancy:

$$P(T_X) = max(P(T_U)/X \in S_X(U_D)) \qquad (11)$$

IV. EXPERIMENTAL RESULTS

A. *Experimental setup*

The algorithm has been evaluated on image sequences from the French LOVe project [12]. The images are taken with a pair of SMAl CMOS cameras, and reduced to quarter VGA resolution 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 occupancy grid computation parameters are set to: $P_{FP} = 0.01$, $P_{FN} = 0.05$, $\tau_O = 0.15$ and $\tau_R = 0.2$.

B. *Results*

Figure 4 shows an overview of the results for 4 different images. For each of the cases (a-d), 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). Recall that in the u-disparity occupancy grid, a vertical line represents the likelihood of occupation throughout a vertical plane of aligned rays of light that reach the focal point.

In case a), the closest obstacle is the truck to the right.

Note that patches of the road are found in various parts of the image, leading to a low probability of occupation in those regions. Other obstacles, such as the car in front of the ego-vehicle, are detected further away, and still are indicated by a very high occupancy likelihood. 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. Also note that some road pixels were found far away, while others were found based on the road markings in front of the truck and stoplight pole in the center. Case c) 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, 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. Note further that a large number of road pixels in the left part of the image increase our confidence that that region is unoccupied. Case d) shows a scene cluttered with pedestrian traffic. In the u-disparity grid, it is obvious that objects are 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 this case remains unknown, and very few road pixels are found. Also, it is clear in these cases that there are strong discretization effects due to the pixel-level sampling and to the estimation of the disparity on integer values. Section V will discuss a method to reduce this effect.

Figure 5 compares this approach to three other methods of using a stereoscopic sensor to create an occupancy grid, using case c from figure 4. Method a) detects the maximum disparity of each column (as in [6]). 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 [11]. As such, much of the image has no information (leading to large areas of 0.5 occupancy probability), and the image that is gained tends to have high certainty. Finally, the method described in [7] is shown in c) and in d) the method detailed in this paper. These methods formally take into account the probability of false measurements, making it less susceptible to noise, and find the obstacle 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 probability values. The inclusion of the road pixels in d) clearly gives a higher likelihood that some areas of the grid are not occupied, keeping the advantages of c) while also performing better in the foreground (like methods a) and b)). Note that when no road pixels are availlable, the results are identical to c).
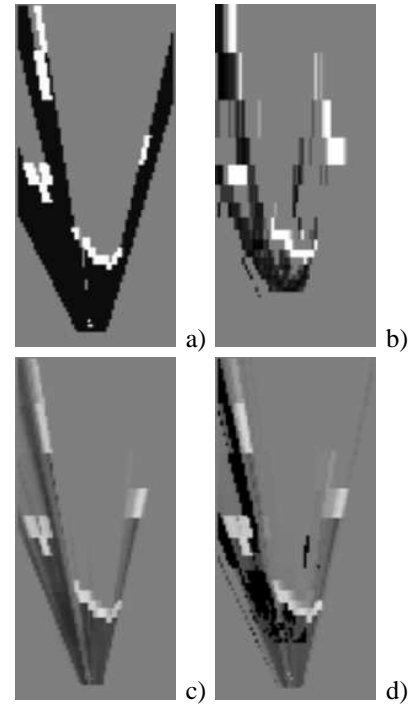

Fig. 5. *Comparison of four approaches to compute the occupancy grid. a) detection of the maximum disparity of each colum, b) exploitation of obstacle and road u-disparity images as measurement point, c) computation in the u-disparity plane, d) computation in the u-disparity plane, with road pixels.*

V. FILTERING THE OCCUPANCY GRID

A. The approach

In order to obtain a smoother and more realistic representation, we propose to apply a filter to the Cartesian occupancy grid. An image-like filter, based on the convolution with a Gaussian kernel, is used. A constant convolution kernel for the complete grid cannot be used, since the uncertainty of measurement is related to the range. Indeed, consdering a constant kernel would require a strong tradeoff when choosing the standard deviation of the filter. A small value would be efficient for close cells, but inaccurate for distant cell, while a large value would suppress many details in the near environnement. We propose to compute a Gaussian filtering kernel H_X for each value of X in the grid.

B. Computation of the convolution kernel

To compute H_X we propose to consider a constant Gaussian filter in the u-disparity plane, and to compute its image through the reconstruction function for each value of X. Thus, in the u-disparity plane, we consider a Gaussian filter H_U, whose variance-convariance matrix is given by:

$$K_U = \begin{bmatrix} \sigma_u^2 & 0.0 \\ 0.0 & \sigma_d^2 \end{bmatrix} \qquad (12)$$

In this model, we consider a standard deviation σ_d^2 along the d axis, which is related to the disparity discretization (e.g. $\sigma_d = 0.5$), and a standard deviation σ_u^2 along the u axis, which is related to the width of the correlation window to model effects like foreground fattening. This is consistant

with the model of the stereoscopic uncertainty proposed in [11].

After linearization of G_D around $U_X = G_D^{-1}(X)$, the image of X in the u-disparity space, the metric convolution kernel can be estimated as:

$$H_X \simeq \mathcal{N}(\mu_X, K_X) \qquad (13)$$

with:

$$\begin{cases} \mu_X &= X \\ K_X &= J_G(U_X).H_U.J_G^T(U_X) \end{cases} \qquad (14)$$

J_G being the Jacobian matrix of G_D. Since the observation function G_D is constant in time, the parameters of the filter for each cell can be computed once and stored to save computation time.

C. Results

Figure 6 shows the effect of the filtering on a previously computed occupancy grid, with the filtered images shown below. These correspond to cases b), c) and d) from Figure 4. The parameters of the filter are $\sigma_d = 0.5$ and $\sigma_u = 2.5$. The image is smoother after the filtering operation, but as expected the details have not been removed. This approach shows particular promise when using a subsequent time filtering or tracking algorithm (*e.g.* the Bayesian Occupancy Filter [3]). Indeed, when driving the vehicle the movements of regions in the occupancy grid are smoother when using the filter. Thus the tracking of cells is made easier.
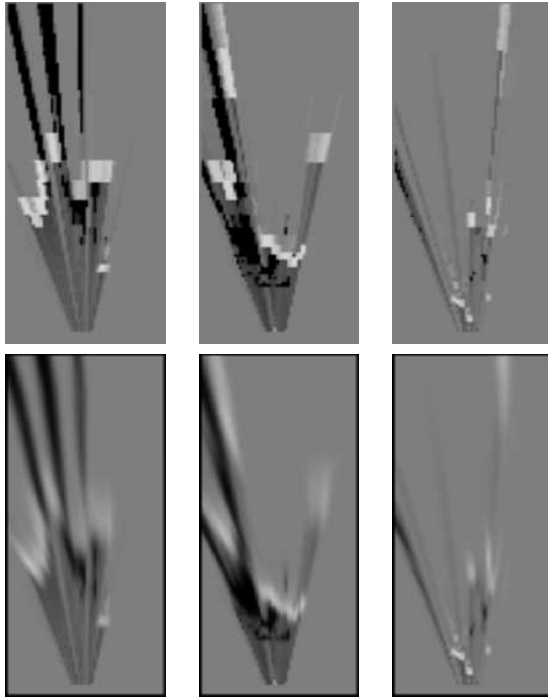

Fig. 6. *Effect of the filter on a metric occupancy grids.*

VI. CONCLUSION AND FUTURE WORK

This paper has presented a novel stereo vision algorithm to create an occupancy grid, which provides four advantages over most other methods. First, real-time operation is possible because using the u-disparity allows computationally efficient and possibly parallel calculations. Second, visibility calculations are simpler in the u-disparity space. Third, 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 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, and uses road pixels' information.

There is considerable work planned to improve this technique. First, we plan to test this algorithm, in conjunction with laser data and Bayesian Fusion, on a new vehicle in the near future. This will provide a second data set for comparison. Second, a GPU-based implementation of this algorithm is expected to further reduce calculation time. Third, the method will be updated to formally consider the semi-occlusions in the stereoscopic images (*i.e.* pixels next to disparity discontinuities). Finally, we hope to refine the approach by considering relationships between system parameters (for example, the probability of a false negative is related to the density of the disparity image).

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," *IJRR*, vol. 25, January 2006.

[4] C. Braillon, C. Pradalier, K. Usher, J. Crowley, and C. Laugier, "Occupancy grids from stereo and optical flow data," in *Proc. ISER*.

[5] L. Matthies and A. Elfes, "Integration of sonar and stereo range data using a grid-based representation," in *IEEE ICRA*, 1988.

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

[7] M. Perrollaz, J.-D. Yoder, A. Spalanzani, and C. Laugier, "Using the disparity space to compute occupancy grids from stereo-vision," in *IEEE IROS*, 2010.

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

[9] R. Labayrade, D. Aubert, and J. Tarel, "Real time obstacles detection on non flat road geometry through v-disparity representation," in *IEEE IV*, 2002.

[10] P. Burt, L. Wixson, and G. Salgian, "Electronically directed "focal" stereo," in *IEEE ICCV*, vol. 0, 1995.

[11] M. Perrollaz, A. Spalanzani, and D. Aubert, "A probabilistic representation of the uncertainty of stereo-vision and its application to obstacle detection," in *IEEE IV*, 2010.

[12] "Love : http://love.univ-bpclermont.fr/."