Probabilistic representation of the uncertainty of stereo-vision and application to obstacle detection

Mathias Perrollaz, Anne Spalanzani And Didier Aubert

***Abstract*— Stereo-vision is extensively used for intelligent vehicles, mainly for obstacle detection, as it provides a large amount of data. Many authors use it as a classical 3D sensor which provides a large tri-dimensional cloud of metric measurements, and apply methods usually designed for other sensors, such as clustering based on a distance. For stereo-vision, the measurement uncertainty is related to the range. For medium to long range, often necessary in the field of intelligent vehicles, this uncertainty has a significant impact, limiting the use of this kind of approaches. On the other hand, some authors consider stereo-vision more like a vision sensor and choose to directly work in the disparity space. This provides the ability to exploit the connectivity of the measurements, but roughly takes into consideration the actual size of the objects.**
In this paper, we propose a probabilistic representation of the specific uncertainty for stereo-vision, which takes advantage of both aspects - distance and disparity. The model is presented and then applied to obstacle detection, using the occupancy grid framework. For this purpose, a computationally-efficient implementation based on the u-disparity approach is given.

I. Introduction

Obstacle detection is a widely explored domain of mobile robotics. It presents a particular interest for the intelligent vehicle community, as it is an essential building block for Advanced Driving Assistance Systems (ADAS). Among the various sensors used for obstacle detection, stereo-vision is very promising, because it provides a rich representation of the scene.

A lot of work deals with stereo-vision for obstacle detection. The main issues addressed are generally the stereo matching process and the conversion of regions of the disparity image into higher level representation [1][2]. For this last aspect, many methods have been proposed. Most of them first transform the stereo data into a large point cloud, and then use it for processing. For example, in [3], the authors apply a clustering algorithm on this cloud, and in [4] it is used through the computation of histograms. Nevertheless, for medium to long range, the data become very sparse, due to the sampling over integer pixel and disparity values. Therefore, the connectivity of points from the same object is lost and algorithms based on proximity may fail. To overcome this problem, before this aggregation step, Nedevschi et al. [2] propose to resample these points to ensure that their density is independent of the range. Other methods use directly the data into the disparity space. This is the case for the u-v-disparity approach [1] or for approaches based on connectivity [5]. These approaches do not take directly into consideration the actual size of objects and often need a post processing stage.

To deal with the specific uncertainty of stereo data, we propose to use a probabilistic approach that takes advantage of both representations. On the one hand, the measurement uncertainty is modelled in the disparity space to consider the specificity of the stereoscopic sensor. On the other hand, the data are used in the metric space, being an easy to use input for the subsequent algorithms. This paper describes this sensor model, proposing three different representations. It also presents an efficient implementation for obstacle detection using occupancy grids. It is not new to use stereo-vision in ocupancy grids [6][7], but our approach gives a different management of uncertainty and a very efficient implementation based on the u-disparity approach.

The paper is organized as follow: in section 2 we describe our probabilistic sensor model for stereo-vision. Three approaches for modeling the sensor uncertainty are proposed. In section 3, a computationaly efficient implementation is given for use in the occupancy grid framework, with a discussion on the computation of the disparity map. In section 4, it is applied to obstacle detection, using the Bayesian Occupancy Filter. Experimental results are given for a real road data set. Finally, section 5 summarizes and discusses future work.

II. The probabilistic sensor model

A. Geometrical developments

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_l) and (u_r, v_r). The length of the stereo baseline is b_s. For clarity, we will consider a coordinate system R_s related to the stereoscopic sensor, and describe the sensor model in it. Extrinsic calibration can be performed to position R_s in any world coordinate system R_w.

Given a point $P(x_s, y_s, z_s)$ in R_s, its position (u_r, d, v) and (u_l, d, v) in the stereoscopic images can be calculated as:

$$\begin{cases} u_r &= u_0 + \alpha_u \frac{x_s - b_s/2}{z_s} \\ u_l &= u_0 + \alpha_u \frac{x_s + b_s/2}{z_s} \\ v &= v_0 + \alpha_v \frac{y_s}{z_s} \\ d &= \alpha_u \frac{b_s}{z_s} \end{cases} \quad (1)$$

where $d = u_l - u_r$ is the disparity value of a given pixel, $v = v_l = v_r$ its y-coordinate. The coordinate system $R_\Delta =$

M. Perrollaz and A. Spalanzani are with the INRIA Grenoble
mathias.perrollaz@inrialpes.fr
D. Aubert is with the LIVIC, INRETS/LCPC,

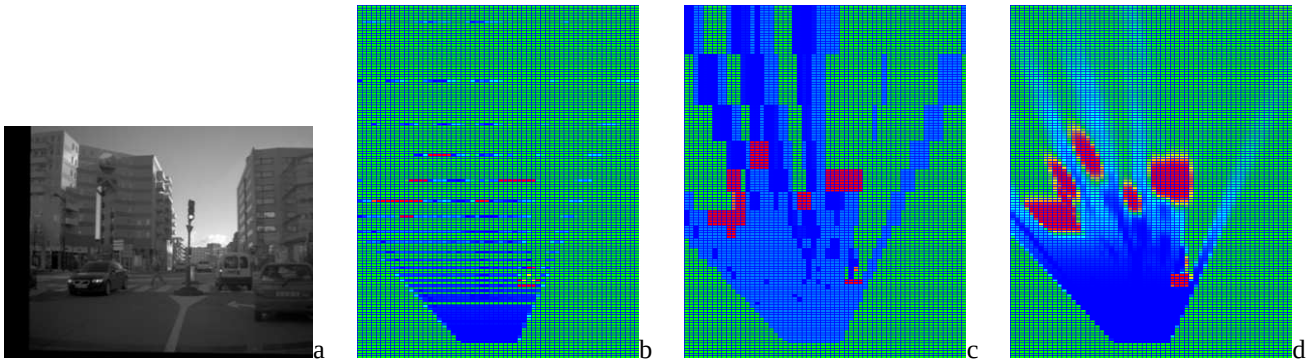

Fig. 1. Left image from the stereo pair, and corresponding observed occupancy grids for the three models. From left to right: punctual, uniform and gaussian. The blue color represents low occupancy probability while the red color represents high occupancy probability. Green means that the probability is 0.5.

$(u = \vec{u_l}, \vec{v}, d)$ defines a 3D space E_Δ, denoted disparity space.
This transform is invertible, so the coordinates in R_s can be retrieved from images coordinates through the reconstruction function G:

$$G : U = (u, v, d)^T \mapsto X = (x_s, y_s, z_s)^T \tag{2}$$

B. The sensor model

In this paper, to model the stereoscopic sensor we will only consider uncertainty due to measurement imprecision. As the measurement is performed through the imaging process and the estimation of the disparity map, it is relevant to model this imprecision in the disparity space. Several approaches can be considered to model this imprecision, considering how the imaging process spreads the probability density over the surface of a pixel. Here we propose three approaches for this modelization, as illustrated with occupancy grids on figure 1.
First, let us consider a point, whose position in R_s is given by random variable X and which is observed by our sensor. We will call Z^Δ and Z^S the random variables representing respectively the measurement in R_Δ and R_s. The probabilistic sensor model consists in defining the distribution $P(Z^S \mid X)$. As we attempt to model the errors in the disparity space, we will first define $P(Z^\Delta \mid X)$, the observation likelihood in R_Δ. Then using transformation G will provide $P(Z^S \mid X)$. Let us also define $\tilde{U}_X$ as the integer coordinates of the image of X in R_Δ.

$$\tilde{U}_X = round(G^{-1}(X)) \tag{3}$$

1) The punctual observation model: First, one could consider that the imaging process attributes all the probability density to the center of a pixel. This approach is equivalent to the classical deterministic transform of the pixels into metric points. The resulting observation model is:

$$P(Z^\Delta \mid X) = \begin{cases} 1 & \text{for } Z^\Delta = \tilde{U}_X \\ 0 & \text{for } Z^\Delta \neq \tilde{U}_X \end{cases} \tag{4}$$

Expressing $P(Z^S \mid X)$ gives a punctual density function centered on $G(\tilde{U}_X)$.

2) The uniform observation model: Second, one can model the error by considering that the projection creates an uniform distribution into the volume $\mathcal{V}_U$ of the voxel centered on $\tilde{U}_X$. This leads to:

$$P(Z^\Delta \mid X) = \mathcal{U}(\mathcal{V}_U) \tag{5}$$

This model leads to:

$$P(Z^S \mid X) = \mathcal{U}(G(\mathcal{V}_U)) \tag{6}$$

3) The gaussian observation model: Finally, for a smoother management of uncertainty, one can model the observation as a gaussian distribution centered on $\tilde{U}_X$.

$$P(Z^\Delta \mid X) = \mathcal{N}(\tilde{U}_X, K_U) \tag{7}$$

We suppose that the errors along u, v and d axis are independent, so:

$$K_U = \begin{bmatrix} \sigma_u^2 & 0.0 & 0.0 \\ 0.0 & \sigma_v^2 & 0.0 \\ 0.0 & 0.0 & \sigma_d^2 \end{bmatrix} \tag{8}$$

To take into account the errors produced by the matching method like foreground fattening, σ_u and σ_v are related to the size of the correlation window, while σ_d only depends on the pixel size ($\sigma_d = 0.5$).
If we linearize G just around the center of the observed pixel $\tilde{U}_X$, $P(Z^S \mid X)$ can be approximated as:

$$P(Z^S \mid X) \simeq N(\mu_D^m, K_D^m) \tag{9}$$

with:

$$\begin{cases} \mu_D^m = G(\tilde{U}_X) \\ K_D^m = J_G(\tilde{U}_X) \cdot K_U \cdot J_G^T(\tilde{U}_X) \end{cases} \tag{10}$$

J_G being the Jacobian matrix of G.

III. Efficient implementation for the occupancy grid framework

A. The u-disparity approach

The u-disparity approach is complement to the v-disparity described in [1]. The idea is to project the pixels of the disparity map along the columns, with accumulation. Then the value of a pixel of coordinates (u, d) in the u-disparity

image is the number of pixels of column u in the disparity image, whose disparity value is d. The resulting image is sort of a bird-eye view representation of the scene, in the disparity space. We will see that it provides an efficient way to implement the computation of an occupancy grid from the stereoscopic data.

B. Road-obstacle separation

For the reminder of the paper, we will consider that we are able to classify pixels from the road surface and pixels from the obstacles. There are several methods to do this, such as estimating the road surface and thresholding the height of the pixels. We will use a double correlation framework to do this, as explained in section III-E.2. After classification, we obtain two u-disparity images, I_U^{obst} and I_U^{road}, respectively containing pixels from the obstacles and from the road surface.

C. Resulting simplification

For occupancy grid computation, we have to consider a detection plane $\mathcal{P}_D$, that 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 to the detection plane. $\vec{x_d}$ is parallel to the baseline and $\vec{y_d}$ is parallel to the optical axis. The coordinates of the center of the baseline are (x_s^o, y_s^o, z_s^o)
For convenience, the pitch, yaw and roll angles of the system will be considered as almost null, so that the detection plane is close to the road surface.
The coordinates in $\mathcal{P}_D$ are:

$$\begin{cases} x_d = & x_S + x_S^o \\ y_d = & y_S + y_S^o \\ z_d = & z_S + z_S^o \end{cases} \tag{11}$$

Considering this constraint over $\mathcal{P}_D$, 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, the easy to compute u-disparity images directly implement the vertical projection on $\mathcal{P}_D$ of the observed points from the scene. The transform between the u-disparity plane $\mathcal{P}_U$ and the detection plane $\mathcal{P}_D$ gives us the simplified observation function G_p:

$$G_p : (u, d) \mapsto (x_d, y_d) \tag{12}$$

With this simplification, the sensor model is reduced to the two dimentions of the grid.

D. Computation of an occupancy grid

1) Accumulating observations from the u-disparity images: As u-disparity images are created by projection, there may be multiple observations for one pixel U. We propose to use an accumulation strategy to take these observations into account. So let us define the total contribution of a pixel U to the X_i point of $\mathcal{P}_D$, as:

$$C_U^X(X_i) = P(Z_S = X_i \mid X) \cdot (I_U^{obst}(U) - I_U^{road}(U)) \tag{13}$$

So the total contribution of the whole set of pixels of the disparity image $\mathbb{U}$ can be expressed as:

$$C_{total}^X(X_i) = \sum_{U \in \mathbb{U}} C_U^X(X_i) \tag{14}$$

A positive contribution expresses a confidence on occupancy for the cell, while a negative contribution means that the cell is not occupied.

2) Occupancy grid computation: To compute the occupancy grid from the observation, the distribution $P(O_i \mid X)$ has to be evaluated. O_i is a boolean random variable, such as $O_i = 1$ means that cell $Cell_i$, centered on X_i, is occupied. To obtain a smooth accumulation, we choose to perform it through a sigmoid function:

$$P(O_i \mid X) = \frac{1}{1 + e^{\frac{C_{total}^X(X_i)}{\sigma}}} \tag{15}$$

When the contribution is equal to zero, $P(O_i \mid X) = 0.5$.

3) Considering the free-field: As explained in [8] for the case of a range finder, if the sensor detects a surface at a certain distance, the field between the sensor and the surface must be empty (except in case of miss-detections). To easily implement such a strategy in stereo-vision, we will not consider each measurements individually, but search for the first detected point for each ensemble of vertically aligned rays. This avoids lowering the occupancy probability of short objects, since many rays go above the object.
The u-disparity representation fits well to this strategy. Each column corresponds to an ensemble of vertically aligned rays. Thus the implementation is straightforeward, since it is sufficient to search the first non-zero pixel in every column of the "obstacle" u-disparity image, starting from the bottom. The free field is then the region between this pixel and the bottom of the column. To take this into account, all the pixels in free field are incremented in the "road" u-disparity image. Figure 2 illustrates this method.

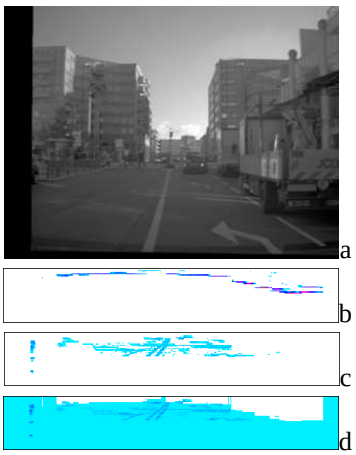

Fig. 2. Left image from the stereo pair (a) and corresponding obstacle (b) and road (c) u-disparity images. The last road u-disparity image (d) is enriched with the free-field information.

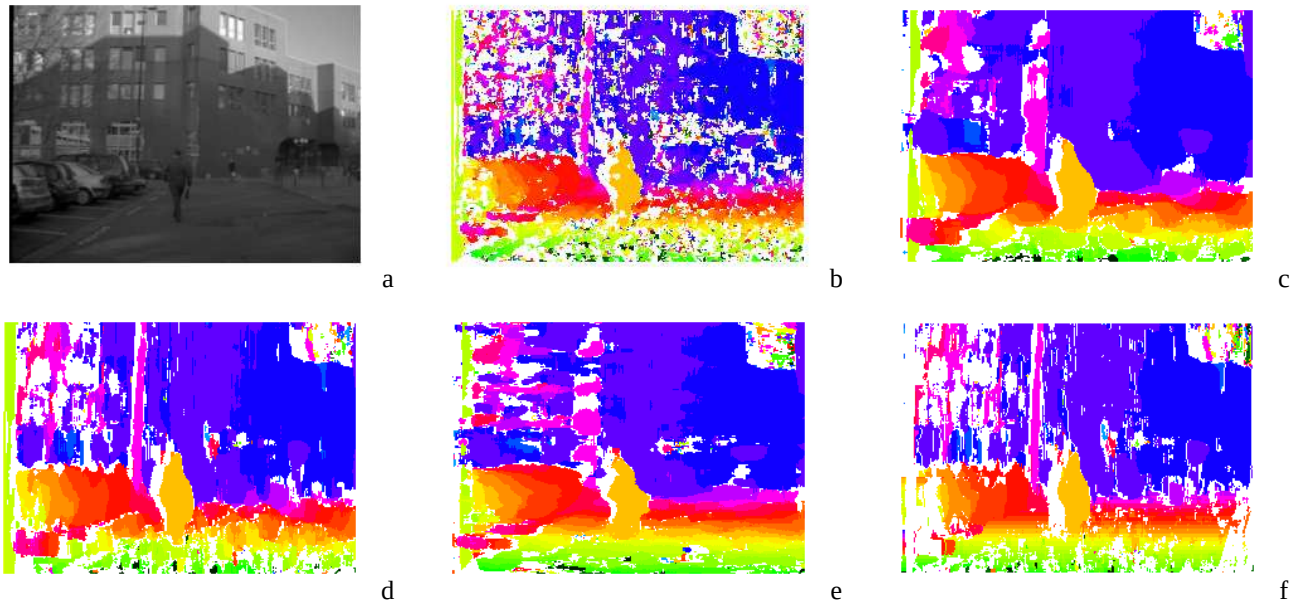

Fig. 3. Disparity maps obtained with various window shapes (width*height)): a) left image of the stereo pair, b) 5*5 pixel, c) 19*19 pixels, d) 7*19 pixels, e) 19*7 pixels with double correlation, f) 7*19 pixels with double correlation.

4) Resulting grid: Figure 1 shows occupancy grids computed with the three different models. The punctual sensor model provides a very sparse representation. On the other hand, the two other models take into consideration that the uncertainty is related to the range. The representation using the gaussian approach is smoother. The free field appears in blue at the bottom of these images.

E. Considerations about the disparity map

Before building an occupancy grid by stereo-vision, it is necessary to compute a disparity map to obtain measurement points. Our approach is based on a classical hypothesis in the field of automotive sensing: the obstacles are vertical, and the road surface (here the detection plane) is horizontal. One can advantageously take this into consideration: the correlation window's size can be adapted to reduce the matching uncertainty.

1) Influence of the window size: Among the large variety of matching algorithms [9], we decided to use a local matching method for the disparity map computation. This is generally the choice made for applications where real time is necessary, due to the relatively low computation cost. This kind of methods supposes that the disparity must be the same over the complete correlation window. This means that a large correlation window leads to errors on objects boundaries (expressed by σ_u and σ_v in equation 8) and to imprecision on the road surface. On the other hand, a small correlation window is not discriminant enough and will produce matching errors. Figures 3-b and 3-c show these behaviors. Many approaches have been proposed to cope with this limitation, such as multiple window [10], weighted window [11] or deformable window [12]. All these methods can lead to significant improvements, but are generally designed for generic applications and suppose an almost squared shape for the correlation window.
On the opposite, one can use the simple hypothesis that the road surface is almost horizontal while obstacles are almost vertical. So vertically, they present an almost constant disparity. As illustrated on figure 3-d, a vertical window reduces the foreground fatening and creates errors on the road surface. On contrary, figure 3-e shows that an horizontal window does well on the road surface (as noticed in [13]), but it creates some errors on objects boundaries.

2) The double correlation: Precisely, knowing the perspective distortion on the road surface, one can adapt the correlation process by applying an homography on one of the stereo images [14], or by using a sheared correlation window [15]. In both cases, the correlation process is performed twice: with classical approach and with the road-compliant approach. These techniques offer better matching capability on the road surface, even with a vertically large correlation window, as illustrated on figure 3-f. Furthermore, they provide direct classification between "road" and "obstacle" pixels: according to which matching process gives better correlation cost, the disparity of a pixel is either reported on a "road" or on an "obstacle" disparity map, without thresholding. In our point of view, such a double correlation paradigm has a third advantage. It allows the use of a mostly vertical correlation window, even for the road surface, with the benefit of a higher precision on boundaries. Thus, in the Gaussian model the value of σ_u can be reduced. The resulting augmentation of the value of σ_v is not a problem, since it has no effect on the occupancy grid implementation.

3) Fast computation: For this approach to be efficient it needs large-size correlation windows, increasing the cost of the aggregation step of the algorithm. We recommend to use

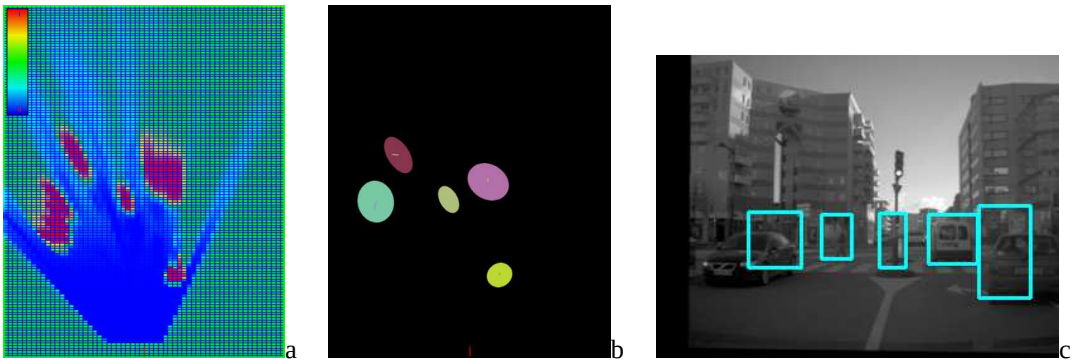

Fig. 4. (a) Estimated grid from the BOF (b) tracks from the FCTA (c) corresponding volumes, projected over the left image of the stereo paire.

integral images for the aggregation step, as proposed in [16]. Aggregation over correlation window is then reduced to one addition and two substractions, whatever the window size.

IV. Application to obstacle detection

A. The Bayesian Occupancy filter

A promising way to deal with the occupancy grids is to use the Bayesian Occupancy Filter (BOF)[17]. The BOF is an adaptation of the Bayesian filtering 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 a velocity grid containing an estimated velocity probability for each cell.

B. The Fast Clustering-Tracking Algorithm

For road applications, it is often necessary to retrieve an object level representation of the scene. This can not be directly reached from the occupancy grid, and therefore a clustering algorithm is necessary. An algorithm adapted to the BOF framework is the "Fast Clustering Tracking Algorithm" described in [18]. It has the interest to create clusters considering not only the connectivity in the occupancy grid, but also in the estimated velocity grid. Thus two connected cells with different speeds are not merged during the clustering process. Figure 4 illustrates the process with the estimated occupancy grid (a), the clustering/tracking (b) and the results projected on the left image (c).

C. Experimental setup

The algorithm has been evaluated on sequences from the french LOVe project, dealing with pedestrian detection. The images are taken with a couple of SMAl CMOS camera, 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 $0m < y_d < 35m$ and the cell size is $0.25m * 0.25m$. The correlation window measures $7 pixels$ in width and $19 pixels$ in height. σ_u is set to the third of the correlation window width and $\sigma_d = 0.5$.
Separately, the matching stage and the detection stage both run in real-time on a laptop, at video framerate.

D. Results

First, the punctual model gives very sparse data, resulting in a very over-segmented decomposition of the scene. In the short range there is enough precision to ensure connectivity between the measurements (compared to the grid resolution) and some detections are correct, but generally it is not suitable for road obstacle detection.
With the uniform model, regions of the grid corresponding to objects generally appear as connected area, so for static scenes the detections results are quite good. Moreover, the filtering capability of the BOF rejects all the errors that appear only on one frame. The main limitation comes with the tracking of mobile objects. With this uniform model, the occupancy probability attached to an object moves roughly as a block. Therefore, the tracking algorithm fails to estimate the velocity correctly.
The Gaussian model corrects this behaviour, since it allows the occupied cells to move smoothly over the grid. As a consequence, the velocities of cells are generally correctly estimated. This is noticeable when the system observes two pedestrians crossing: the velocity direction of their respective group of cells being very different, they are not agglomerated in the clustering stage.
The results obtained with the Gaussian approach on real data are illustrated on figure 5. Note that FCTA does not provide the height of objects, so for visualization the height of bounding boxes is arbitrarily set to $1.8m$. Results are very promising, since most of the obstacles are correctly detected, segmented and tracked. Only the lack of texture on certain objects lead to detection failures. The limit of the tracking capability of the BOF appears with objects which have a high lateral velocity. For them, the estimated position is often a bit late, due to the constant velocity model used in the FCTA.

V. Conclusion and future work

Uncertainty is a problem when using stereo-vision for 3D measurements, since the imprecision is directly linked to the distance of observed object. We proposed in this paper three possible probabilistic representations of the uncertainty. Particularly, our gaussian model has proven to be the most efficient, as it provides a smoother representation. This makes the clustering and tracking tasks easier and more robust.

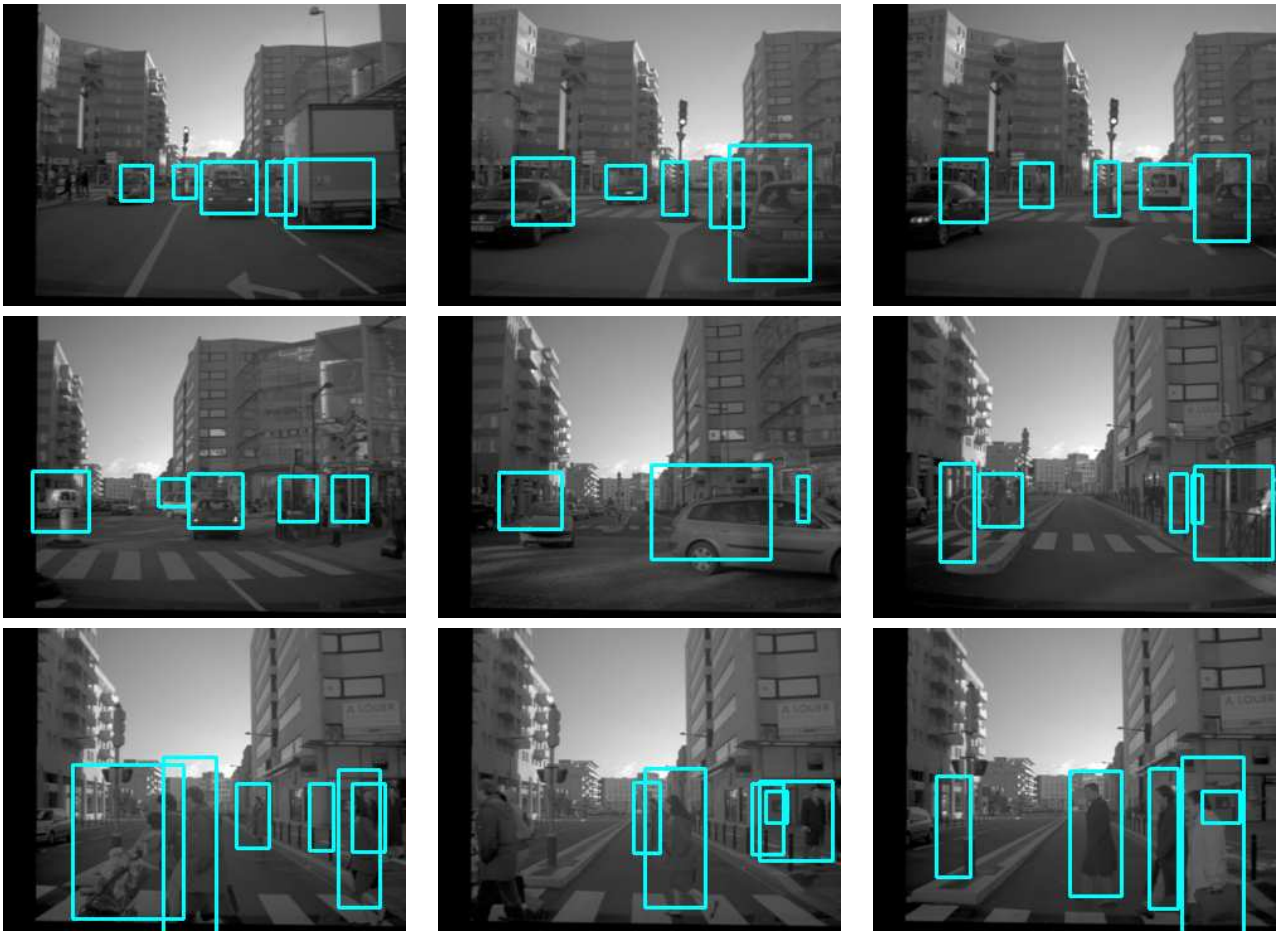

Fig. 5. Results obtained for a sequence of stereo data from the LOVe project.

Moreover, we proposed an implementation of our approach in the Bayesian Occupancy Filter framework. The use of the u-disparity representation helps in reducing the computational complexity and the method showed very promising results on a real road data set. Considering that this framework is also well suited for sensor fusion, since it performs fusion at the grid cell level, we plan to use it for fusion between stereo-vision and range-finder data.

ACKNOWLEDGMENT

The authors would like to thank the French projects LOVe[19] and Arosdyn[20] for funding this research.

REFERENCES

[1] R. Labayrade, D. Aubert, and J. Tarel, "Real time obstacles detection on non flat road geometry through v-disparity representation," in *IV.*, 2002.
[2] 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 *IES.*, 2004.
[3] U. Franke and A. Joos, "Real-time stereo vision for urban traffic scene understanding," in *IV*, 2000.
[4] V. Lemonde and M. Devy, "Obstacle detection with stereovision." in *Mechatronics and robotics*, 2004.
[5] T. Veit, "Connexity based fronto-parallel plane detection for stereovision obstacle segmentation," in *ICRA*, 2009.
[6] L. Matthies and A. Elfes, "Integration of sonar and stereo range data using a grid-based representation," in *ICRA*, 1988.
[7] C. Braillon, C. Pradalier, K. Usher, J. Crowley, and C. Laugier, "Occupancy grids from stereo and optical flow data," in *ISER*, 2006.
[8] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.
[9] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, 2002.
[10] A. Fusiello, V. Roberto, and E. Trucco, "Efficient stereo with multiple windowing." in *CVPR*, 1997.
[11] K. Yoon and I. Kweon, "Locally adaptative support-weight approach for visual correspondence search." in *CVPR*, 2005.
[12] O. Veksler, "Stereo correspondence with compact windows via minimum ratio cycle." *PAMI*, vol. 24(12), 2002.
[13] S. Lefevbre, S. Ambellouis, and F. Cabestaing, "Obstacle detection on a road by dense stereovision with 1d correlation windows and fuzzy filtering." in *ITSC.*, 2006.
[14] T. Williamson, "A high-performance stereo vision system for obstacle detection," Ph.D. dissertation, CMU, 1998.
[15] M. Perrollaz, R. Labayrade, R. Gallen, and D. Aubert, "A three resolution framework for reliable road obstacle detection using stereovision," in *MVA Conf.*, 2007.
[16] O. Veksler, "Fast variable window for stereo correspondence using integral images," in *CVPR*, 2003.
[17] C. Coue, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessiere, "Bayesian occupancy filtering for multi-target tracking: an automotive application." *IJRR*, vol. 25, 2006.
[18] K. Mekhnacha, Y. Mao, D. Raulo, and C. Laugier, "Bayesian occupancy filter based "Fast Clustering-Tracking" algorithm," in *IROS*, 2008.
[19] "Love : http://love.univ-bpclermont.fr/."
[20] "Arosdyn : http://arosdyn.gforge.inria.fr/."