



HAL
open science

Obstacle Detection Using a Single Camera Stereo Sensor

Luc Duvieubourg, Sébastien Ambellouis, Sébastien Lefebvre, François Cabestaing

► **To cite this version:**

Luc Duvieubourg, Sébastien Ambellouis, Sébastien Lefebvre, François Cabestaing. Obstacle Detection Using a Single Camera Stereo Sensor. Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, SITIS'07, Dec 2007, Shanghai, China. pp.979-986, 10.1109/SITIS.2007.26 . hal-00521115

HAL Id: hal-00521115

<https://hal.science/hal-00521115>

Submitted on 25 Sep 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.

Obstacle detection using a single camera stereo sensor

Luc Duviembourg¹, Sébastien Ambellouis², Sébastien Lefebvre² and François Cabestaing¹

¹LAGIS - UMR CNRS 8146
Bâtiment P2 - USTL, Cité Scientifique
59655 Villeneuve d'Ascq, FRANCE
luc.duviembourg@univ-lille1.fr,
fcab@ieee.org

²INRETS-LEOST,
20, rue Elisée Reclus
59666 Villeneuve d'Ascq, FRANCE
sebastien.ambellouis@inrets.fr,
sebastien.lefebvre@inrets.fr

Abstract— In this paper we present a catadioptric stereovision system based on a single camera, two plane mirrors and a prism mounted on a rotating axis. On the one hand, the system is able to project a real scene on two half images. Thus, contrary to classical stereo system, it deals with the problem of the camera synchronisation and makes the calibration and rectification process easier. On the other hand, the optical axis of the system can be steered to point a region of interest of the real scene. This new system is cheaper and more compact than two cameras based setup. It has been developed in the context of on-road vehicle applications and more specifically for long-range road surveillance. This degree of liberty is useful when the region of interest is far away in the front of the vehicle because it requires a long focal length to reach a sufficient resolution. This optical system is associated to a dense stereo matching algorithm and a disparity map segmentation process. We present some results we have obtained using synthetic stereo images to illustrate the functionality of this setup.

I. INTRODUCTION

Obstacle detection process is a critical task in the context of development of a driving assistance system. Researchers have investigated this topic for many years. Actually, the most efficient systems are based on several sensors as cameras, lidars or radars. These multi-sensor architectures take into account redundant and complementary information to make the object detection and tracking more accurate and robust.

To guarantee the best efficiency of the system, obstacles have to be detected as far as possible, especially when the speed of the vehicle is high. For radars and lidars, the range of perception is up to 100 meters. For cameras, it is more difficult to provide an efficient surveillance beyond 50 meters because of a trade-off between resolution and field of view. When the vehicle moves on a straight road (cf. figure 1.a), a long focal lens can be mounted on the camera to simultaneously provide sufficient resolution and capture all the width of the road. When turns occur, as shown on figures 1.b, 1.c and 1.d, to keep in sight the region of interest (i.e. the road and its shoulders) located at a greater distance, a wide field of view is required. In this case the image resolution is too low and we have to face to a less accurate computed distance and to a worse detection capability. Moreover, for narrow field of view, some objects in the scene will not be seen, as illustrated on figure 1.c for the pedestrian and the cyclist.

In this paper, we propose a new stereovision system that aims at tracking the region of interest to keep it in sight and computing accurate depth map as shown in figure 1.d. The system is a catadioptric setup that comprises two fixed mirrors, a movable prism and a single camera. The use of only one video sensor makes the system more compact and less expensive. In addition, to steer its field of view is easier than for a classical stereo sensor that requires to move both cameras. Stereo matching performance is improved because the difference between the focal lengths of the two cameras disappears and the intensities differences between stereo pair are reduced. Thus it is easier to calibrate the system.

During the last years, the computer vision community has shown an increasing interest in dioptric and catadioptric stereo systems based on a single camera. Baker and Nayar [1] were the first to study the image formation with catadioptric sensor. In [2] the authors describe a first system based on planar mirrors that project a rectified stereo pair on a the camera sensor. In [3], different camera-mirrors configurations are studied, emphasized on criteria like accuracy of depth reconstruction, field of view and resolution. They proposed four configurations whose two have a central mirror. In [4] and [5] the authors use a biprism in the place of the planar mirror.

The system we propose is an adaptation of the first one proposed by Innaba [6] that has been efficiently implemented by Mathieu and Devernay by taking into account many technological constraints [7].

The structure of the paper is the following. In section II, we describe briefly the architecture of our sensor and we analyse more specifically the behavior of the sensor when a rotation is applied to the prism around its vertical axis. In section III, we present the experimental setup we have designed for obstacle detection application. In section IV, we present some results on disparity map computation and segmentation in a synthesis image context. The conclusions and perspectives are provided in the last section.

II. THE ORIENTABLE SENSOR CONCEPT

A. System description

A schematic top view of the stereoscopic sensor is presented in figure 2. It is composed of two lateral planar mirrors ('(a) and '(b) in figure 2) and of a central prism

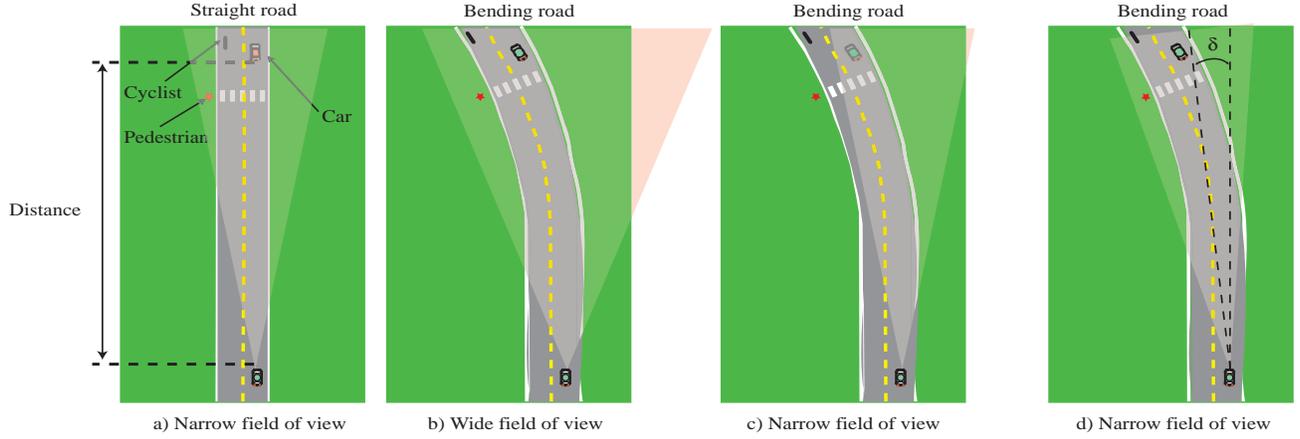


Fig. 1. Wide and narrow field of view

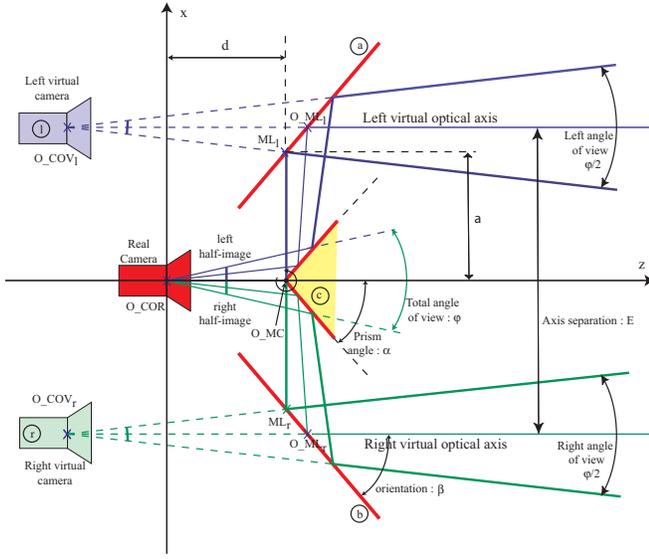


Fig. 2. Schematic top view

with two planar reflective surfaces (marked as '(c)'). This setup is similar to the system described in [7], except that the prism can turn around the axis defined by the edge at the intersection of the reflective surfaces (marked as '(O_MC)'). The optical axis of the camera intersects the edge of the prism, which is projected as a vertical straight line through the center of the image. This straight line, which does not move when the prism is rotated, splits the single image into two half images that would be separately formed by the virtual cameras marked as '(l)' and '(r)' in figure 2.

Let φ denote the total angle of view of the camera, α the angle between the reflective surfaces of the prism, and β the angle between a lateral mirror and the optical axis

of the camera. Assuming a pinhole model, φ is equal to $2 \cdot \arctg(w/(2 \cdot f))$, where w is the width of the image sensor and f the focal length. The angle of view of both virtual cameras is only one half of φ . It can be shown that the optical axis of a virtual camera is parallel to the optical axis of the real camera if β is equal to $\alpha - \varphi/8$. This configuration simplifies the triangulation computations performed during 3D reconstruction, since the optical axes of the virtual cameras are parallel. The lateral mirrors are spaced in order to get a predefined distance E between the optical axes. For simplicity, but an exact calculation could naturally be used, we assume that φ is small enough to be considered equal to its tangent. Under this assumption, two parallel light rays separated by E will reach two pixels with the same coordinates in each half image. Our central prism is mounted on a motorized stage whose angular position can be precisely oriented thanks to a stepper motor. The geometric analysis of the rotation of the prism follows.

B. Prism rotation analysis

As illustrated on figure 3 when a rotation is applied to the prism, the orientation of the optical axes of the virtual cameras change but remain parallel.

The straight line joining the optical centers of virtual cameras is no more perpendicular to the virtual optical axes, but remains perpendicular to the optical axis of the real camera.

Let d denote the distance between the optical center and the edge of the prism, and δ the angle between the symmetry axis of the prism and the optical axis of the real camera. The optical center of each virtual camera remains on a circle of radius d centered on a point that does not belong to the optical axis. We can demonstrate that both optical centers belong to the straight line that is the superior or inferior

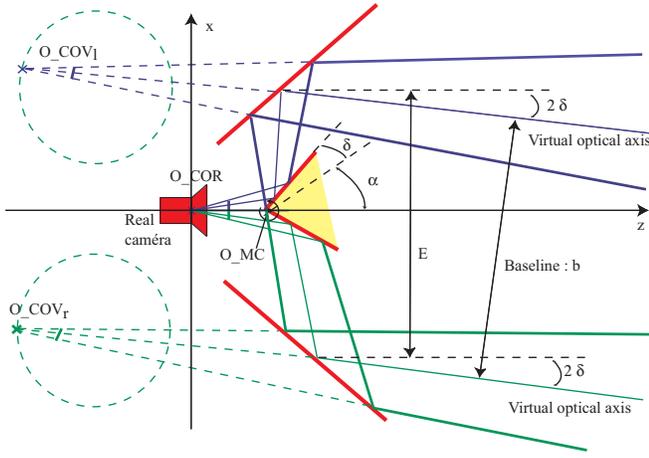


Fig. 3. Stereoscopic setup with rotation of the prism

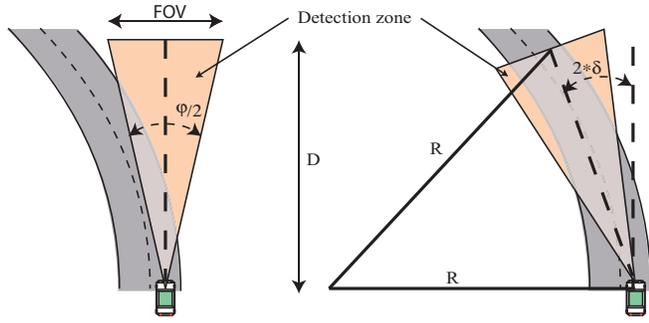


Fig. 4. Curvature radius

limit of the field of view of respectively the left and the right virtual camera.

The angle between the optical axis of a virtual camera and the optical axis of the real one is twice the angle δ . The baseline of our stereo system is the distance between the two virtual parallel optical axes. Its length b is given by the following equation :

$$b = E \cdot \cos(2 \cdot \delta) \quad (1)$$

For very small values of δ , the baseline is equal to the baseline of the system when $\delta = 0$. For high value of δ , the real value of the baseline b has to be taken into account for depth computation.

The next table illustrates how δ varies if we aim at always tracking the road as shown on figure 4. The rotation angle δ of the central prism is evaluated by the simple relation of equation 2.

$$\delta = \frac{1}{2} \cdot \text{asin}(D/(2 \cdot R)) \quad (2)$$

where R is the curvature radius and D the perception distance. The table I gives 3 examples with realistic curvature radii, found in main road and highway specifications for a detection distance of $D = 100m$ and for $E = 0.2m$. With these results we can verify that the variations of the baseline value are not significant to be taken into account for depth computation in the context of our application.

TABLE I
ROTATION ANGLE AND BASELINE

Curvature radius (m)	500	1000	1500
Prism rotation $\delta(^{\circ})$	2.8696	1.4330	0.9551
Baseline b (m)	0.19975	0.19994	0.19997

III. THE STEREOVISION SYSTEM

Let E denote the baseline of the virtual stereovision system, d the distance between the edge of the prism and the CCD sensor of the camera and β the orientation of the lateral mirrors. It is possible to compute the position of the lateral mirrors and the coordinates of the optical centers of the virtual cameras according to these parameters.

The position of lateral mirrors are defined by the coordinates of ML_l and ML_r given by the following equations :

$$\begin{cases} x_{MLl} = a \\ y_{MLl} = 0 \\ z_{MLl} = d \end{cases} \quad \begin{cases} x_{MLr} = -a \\ y_{MLr} = 0 \\ z_{MLr} = d \end{cases} \quad (3)$$

With :

$$a = \frac{E - 2 \cdot d \cdot \cos(2 \cdot \beta)}{2(1 + \cos(2 \cdot \beta))} \quad (4)$$

The coordinates of the left and the right virtual optical centers O_{COV_l} and O_{COV_r} are calculated by the following relations :

$$\begin{cases} x_{COVl} = a - (a + d) \cdot \cos(\pi - 2 \cdot \beta) \\ y_{COVl} = 0 \\ z_{COVl} = d - (a + d) \cdot \sin(\pi - 2 \cdot \beta) \end{cases} \quad (5)$$

$$\begin{cases} x_{COVr} = -a + (a + d) \cdot \cos(\pi - 2 \cdot \beta) \\ y_{COVr} = 0 \\ z_{COVr} = d - (a + d) \cdot \sin(\pi - 2 \cdot \beta) \end{cases} \quad (6)$$

Our experimental system has been designed after we have defined all the parameters available in the table II. We define a field of view so that the road and its shoulders are detected at a distance of $100m$. These parameters have been chosen to respect a trade-off between the compactness and the accuracy of the system.

Table III gathers the geometric parameters of the system computed from previous information.

TABLE II
FIXED PARAMETERS

f	35mm	Focal distance
w	8.9mm	CCD sensor width
D	100m	Distance
E	0.20m	Axis separation
d	0.10m	Distance between optical center and the prism
α	45°	Angle between prism and optical axis

Finally, Table IV shows the optical center coordinates of the virtual camera.

TABLE III
SENSOR PARAMETERS

FOV	$12.71m$	FOV of the sensor
β	43.1875°	Angle between lateral mirror and optical axis
O_{MC}	$(0; 0; d)^T$	Prism origin
ML_l	$(0.0881; 0; d)^T$	Left lateral mirror position
ML_r	$(-0.0881; 0; d)^T$	Right lateral mirror position
O_{COR}	$(0; 0; 0)^T$	Real optical center

TABLE IV
VIRTUAL OPTICAL CENTER POSITION

O_{COV_l}	$(E/2; 0; -0.0877)^T$	Left optical center origin
O_{COV_r}	$(-E/2; 0; -0.0877)^T$	Right optical center origin

IV. THE SIMULATION RESULTS

A laboratory prototype has been constructed with mirrors carefully mounted on an optic table (cf. figure 13). We are now developing the mechanical adjustments in order to obtain usable images. Since an embedded version of this prototype is still under construction, range image sequences of real scenes are not yet available. To provide a quantitative evaluation of the improvements brought by the proposed sensor, synthetic stereo images were generated thanks to a graphical simulation software package i.e. *POVRAYTM* [8]. The proposed sensor has been designed using this tool. Classical lane configuration with real world textures for the road surface and for horizontal curvatures of $1000m$ were used. With these considerations we generate images of resolution 1024×512 pixels. These images of the scene content, a blue car at $60m$, a green car and a pedestrian (red cylinder) at $100m$ and an orange car at $200m$. Figures 5 and 6 show stereo images of the scene obtained by the virtual sensor respectively by a lens with focal length equal to $16mm$ and $35mm$.

Figure 7 shows that, without rotation of the prism the pedestrian and two cars do not appear. To perceive again the pedestrian and the cars located at $100m$, it is necessary to apply a rotation of 1.5° . As illustrated in the figure 8, to place the most distant car and the pedestrian at the center of the field of view, the prism rotation is 2.5° . The figure 9 shows that it is possible to look at the nearest vehicle if the

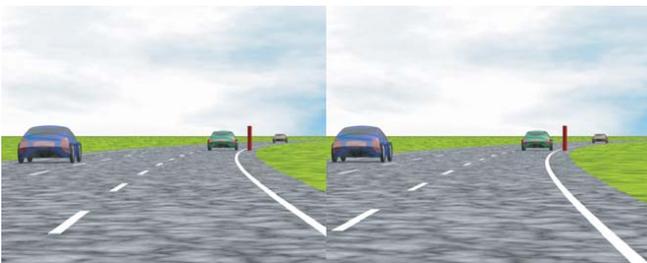


Fig. 5. The observed scene with $f = 16mm$ and $\delta = 0$



Fig. 6. The scene with $f = 35mm$ and $\delta = 0$

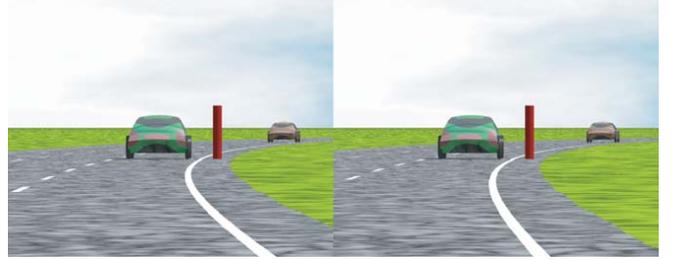


Fig. 7. The scene with $f = 35mm$ and $\delta = +1.5^\circ$

prism rotation is -3° . In all the previous cases, more details of each vehicle are available on the images.

A. Dense disparity map computation

All these images have been rectified in order to apply the dense stereo matching algorithm presented in [9]. This algorithm, named Similarity Based Adaptive Neighborhood (henceforth SBAN), is based on a SAD measure and excludes the pixels of the neighborhood that are not similar to its center. It corresponds to adjusting the size and the shape of the correlation neighborhood according to the local color or intensity information.

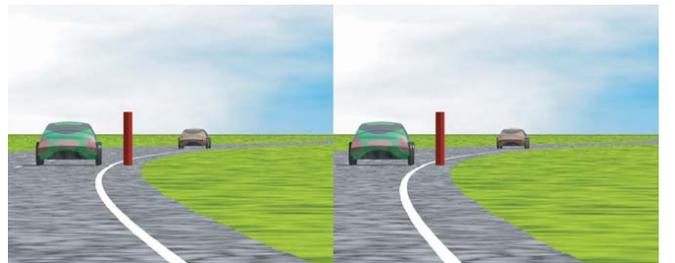


Fig. 8. The scene with $f = 35mm$ and $\delta = +2.5^\circ$

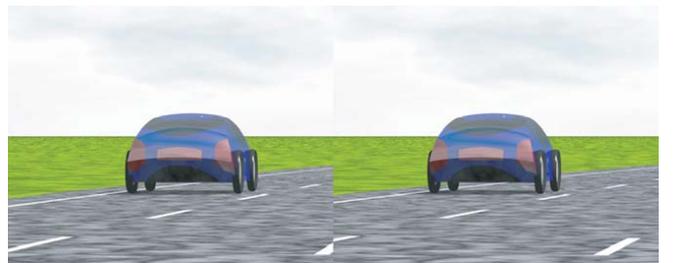


Fig. 9. The scene with $f = 35mm$ and $\delta = -3^\circ$



Fig. 10. Disparity map of image pair of the figure 7

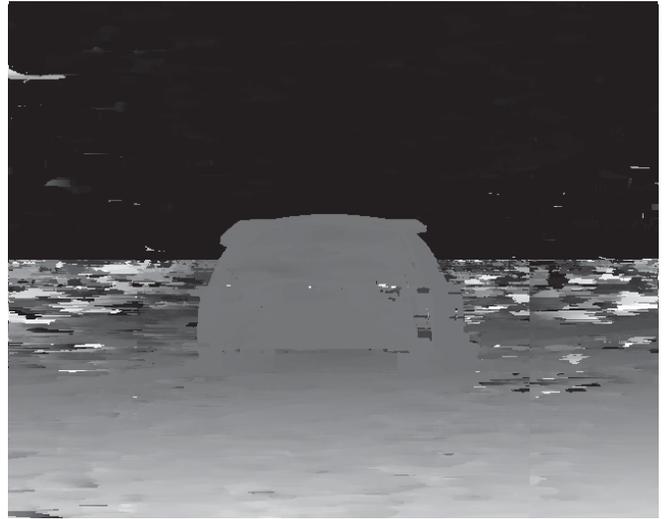


Fig. 12. Disparity map of image pair of the figure 9



Fig. 11. Disparity map of image pair of the figure 8

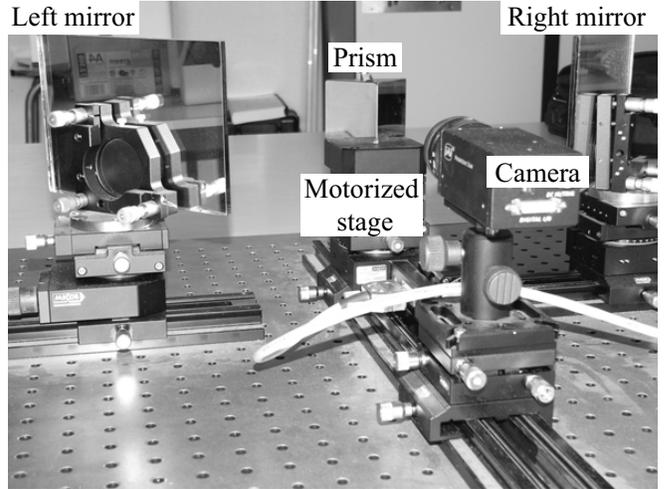


Fig. 13. Laboratory prototype

The computed range images obtained respectively from the images 7, 8 and 9 are shown in figures 10, 11 and 12. The highest gray levels (white pixels) in these range images correspond to close objects and the lowest gray levels (black pixels) to distant objects. Only the depth values of each right images have been computed.

B. Disparity map segmentation

Road and obstacle are detected by analysing the disparity map in two steps. The first step defines the set of pixels corresponding respectively to the vehicles and to the road. The second step yields the detection and the separation of each obstacle in front of the stereo sensor.

In the literature, authors generally firstly compute the ground plane parameters by fitting a plane in the (x, y, d) space using a least squares method. Secondly, they verify that each pixel of the disparity map satisfies inequality (8). In [10], it is considered that a road and an obstacle can be

approximated respectively by an horizontal and a vertical plane in the real scene. As it is described in [10], road disparity $d_{road}(x, y)$ is a linear function in terms of disparity map pixels coordinates:

$$d_{road}(x, y) = ax + by + c, \quad (7)$$

where (a, b, c) are the parameters of the ground plane and (x, y) are pixel coordinates. Under this assumption, a pixel corresponds to an obstacle if its disparity value satisfies the following equation:

$$d(x, y) > ax + by + c. \quad (8)$$

In [11], Labayrade *et al.* assume the same hypothesis but in the histogram of disparity values space. This is a robust method for detecting the obstacles which is based on the assumption that, for each scanline where the road is visible, the dominant disparity value is that of road surface pixels. A "V-disparity" image is computed for each scanline. A "V-disparity" is the histogram of disparity values



Fig. 14. Disparity map of the stereo pair of the figure 7 with bounding boxes



Fig. 15. Disparity map of the stereo pair of the figure 8 with bounding boxes

of the associated scanline in the original disparity image. The ground plane is estimated by extracting from this "V-disparity" image the dominant line features with the Hough transform.

In this work, we remain in the (x, y, d) space and we propose to analyse separately the profile of every column in the disparity map [12]. For each column x of the disparity map, the disparity profile of road pixels is a straight line segment with a large slope. On the contrary, for the pixels corresponding to an obstacle, the slope of the straight line is small. Thus, boundaries of an obstacle are marked at the intersections of two line segments in the profile, one with a small slope and one with a large slope. Bounding boxes are generated using the edges detected on neighbour columns with the constraint of constant disparity to define each bounding box.

The results are presented on the figures 14, 15, 16 for which one can visualise the bounding boxes superimposed on the right range image. Each obstacle is detected. More specifically, figures 14 and 15 show clearly that the most distance vehicle is detected with accuracy and certainty thanks to the orientation of the optical axis of the sensor. The quality of the detection would have been worse using the stereo pair of the figure 5.

V. CONCLUSIONS

The work, presented in this paper, has been done in the framework of the Ravioli's project (RAdar, VISION Orientable, LIdar). This project deals with the development of a multi sensor architecture including Radar, Lidar and video sensors in the context of long range obstacle detection and tracking.

In this paper, we have presented a new catadioptric stereo sensor developed in the project. It is based on a single camera with a long focal length lens and used a movable prism to steer the optical axis of the sensor toward a given region of interest. Firstly, the system yields a good resolution of



Fig. 16. Disparity map of the stereo pair of the figure 9 with bounding box

each object on the road. Secondly, it offers a complete view of the scene. We have validate its fonctionnality thanks to a disparity map computation and segmentation process. The results show that the sensor seems a valuable concept for applications that require an accurate 3D object reconstruction, description and tracking.

Now, we are working on an embedded prototype to carry out experiments on real world scenes. This sensor will be connected to a FPGA based architecture hardware, called the STREAM processor, that we have developed in a previous project. To reach a real-time obstacle detection application, the adequation between the algorithm implementation and the computing architecture is required. Thus, we aim at implementing the rectification stage, the disparity map computation and segmentation algorithm on the STREAM. We will improve obstacle detection and tracking by using a lane-markings extraction process.

VI. ACKNOWLEDGMENTS

This work is part of the french research project RaViOLi, supported by EU, french government and the Nord-Pas-de-Calais regional council under contract number OBJ2-2005/3-4.1-253-7820.

REFERENCES

- [1] S. Baker and S. K. Nayar, "A theory of catadioptric image formation," in *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 1998, p. 35.
- [2] J. Gluckman and S. Nayar, "Rectified Catadioptric Stereo Sensors," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, Jun 2000, pp. 380–387.
- [3] E. Mouaddib, R. Sagawa, T. Echigo, and Y. Yagi, "Stereo vision with a single camera and multiple mirrors," in *Proc. IEEE International Conference on Robotics and Automation*, Barcelona, Spain, April 18–22 2005, pp. 812–817.
- [4] D. H. Lee, I.-S. Kweon, and R. Cipolla, "A biprism-stereo camera system," in *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR'99*. IEEE Computer Society, 1999, pp. 1082–1087. [Online]. Available: <http://dblp.uni-trier.de/db/conf/cvpr/cvpr1999.htmlLeeKC99>
- [5] D. H. Lee and I. S. Kweon, "A novel stereo camera system by a biprism," *IEEE Trans. on Robotics and Automation*, vol. 16, no. 5, pp. 528–541, Oct. 2000.
- [6] M. Innaba, T. Hara, and H. Inoue, "A stereo viewer based on a single camera with view-control mechanisms," in *Int. Conf. on Intelligent Robots and Systems*, Yokoama, Japon, July 1993.
- [7] H. Mathieu and F. Devernay, "Système de miroirs pour la stéréoscopie," INRIA, Sophia Antipolis, Rapport technique 172, Juin 1995, projet Robotvis.
- [8] POVRAY, "The persistence of vision raytracer (PovRay)," <http://www.povray.org/http://www.povray.org/>. [Online]. Available: <http://www.povray.org/>
- [9] M. Pérez-Patricio, O. Colot, and F. Cabestaing, "A SBAN stereovision algorithm using hue as pixel similarity criterion," in *International Conference on Computer Vision and Graphics, ICCVG'2004. Published in special issue of Computational Imaging and Vision, Vol. 32, Dec. 2005*, K. Wojciechowski, B. Smolka, H. Palus, R. Kozera, W. Skarbek, and L. Noakes, Eds. Warsaw, Poland: Springer, Sept. 2004. [Online]. Available: [perez-iccvg-2004.pdf](#)
- [10] S. Se and M. Brady, "Ground plane estimation, error analysis and applications," *Robotics and Autonomous Systems*, no. 39, pp. 59–71, 2002.
- [11] R. Labayrade, D. Aubert, and J. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through V-disparity representation," in *Procs. IEEE Intelligent Vehicles Symposium 2002*, Versailles, France, June 2002, pp. 646–651.
- [12] S. Lefebvre, S. Ambellouis, and F. Cabestaing, "Obstacles detection on a road by dense stereovision with 1D correlation windows and fuzzy filtering," in *IEEE International Transportation Systems Conference*, Toronto, Canada, Sept. 2006, pp. 739–744.