

Real Time Obstacle Detection in Stereovision on Non Flat Road Geometry Through "V-disparity" Representation.

Raphael Labayrade, Didier Aubert, Jean-Philippe Tarel

Abstract— Many roads are not totally planar and often present hills and valleys because of environment topography. Nevertheless, the majority of existing techniques for road obstacle detection using stereovision assumes that the road is planar. This can cause several issues : imprecision as regards the real position of obstacles as well as false obstacle detection or obstacle detection failures. In order to increase the reliability of the obstacle detection process, this paper proposes an original, fast and robust method for detecting the obstacles without using the flat-earth geometry assumption; this method is able to cope with uphill and downhill gradients as well as dynamic pitching of the vehicle. Our approach is based on the construction and investigation of the "v-disparity"¹ image which provides a good representation of the geometric content of the road scene. The advantage of this image is that it provides semi-global matching and is able to perform robust obstacle detection even in the case of partial occlusion or errors committed during the matching process. Furthermore, this detection is performed without any explicit extraction of coherent structures such as road edges or lane-markings in the stereo image pair. This paper begins by explaining the construction of the "v-disparity" image and by describing its main properties. On the basis of this image, we then describe a robust method for road obstacle detection in the context of flat and non flat road geometry, including estimation of the relative height and pitching of the stereo sensor with respect to the road surface. The longitudinal profile of the road is estimated and the objects located above the road surface are then extracted as potential obstacles; subsequently, the accurate detection of road obstacles, in particular the position of tyre-road contact points is computed in a precise manner. The whole process is performed at frame rate with a current-day PC. Our experimental findings and comparisons with the results obtained using a flat geometry hypothesis show the benefits of our approach. Future work will be concerned with the construction of a 3D road model and the test of the system for Stop'n'Go applications.

Keywords— imaging and vision enhancement, stereoscopic vision, road obstacle detection, non flat road geometry, semi-global matching, real time processing.

I. INTRODUCTION

IN the field of stereo vision-based road obstacle detection, a number of assumptions are frequently made about the environment in order to facilitate the process. The majority of existing techniques assumes that the road is planar [1][2]. Some techniques assume that there is a constant plane [3], others perform a dynamic estimate of the inclination of the road plane and the relative height of the stereoscopic sensor [4][5]. In [4], lane-markings are extracted and matched between the left and right images in order to obtain the parameters of the road plane; however this method is limited to the precision of lane-markings extraction tech-

niques which are sometimes not reliable (poor quality of lane-markings, etc.). Another possibility is to model the suspensions of the vehicle; however the imperfections of the road surface (holes etc.) can not be taken into consideration. This paper presents a method for robustly estimating the road parameters -including vehicle pitching- which is not affected by these problems. Furthermore, while the hypothesis of a flat road geometry is reasonable in the vicinity of the vehicle, it may not be valid for the entire part of the road visible in the image. Obstacle detection is therefore only reliable on the locally planar area, which limits the detection distance. In order to cope with non flat geometry roads, this paper presents a robust method for locally estimating the longitudinal road profile. Our method does not need any extraction of lane-markings but exploits all the relevant information in the image (texture of the road, shadows, road edges, etc.). Thanks to this estimation, the result of the obstacle detection process is far more reliable and accurate.

Our approach is based on the construction and subsequent processing of the "v-disparity" image, which provides a good and robust representation of the geometric content of road scenes.

This paper is structured as follows: Section II presents the models we have used with respect to the stereoscopic sensor and the longitudinal road profile. Section III explains the construction of the "v-disparity" image and describes its main properties. Section IV presents our method for performing obstacle detection. Experimental results are shown including comparisons with the results obtained using a flat road geometry assumption. Lastly, Section V concludes the paper and discusses the potential for future work.

II. ASSUMPTIONS AND MODELS

A. Modelling the stereo sensor

The two image planes of the stereo sensor are supposed to belong merely to the same plane and are at the same height above the road (see Fig. 1). This camera geometry means that the epipolar lines are parallel.

In what follows we will need to perform positioning in three coordinate systems shown in Fig. 1: R_a (absolute), R_{cr} (right camera) and R_{cl} (left camera). R_a is the road coordinate system.

¹v is the ordinate of a pixel in the (u,v) image coordinate system

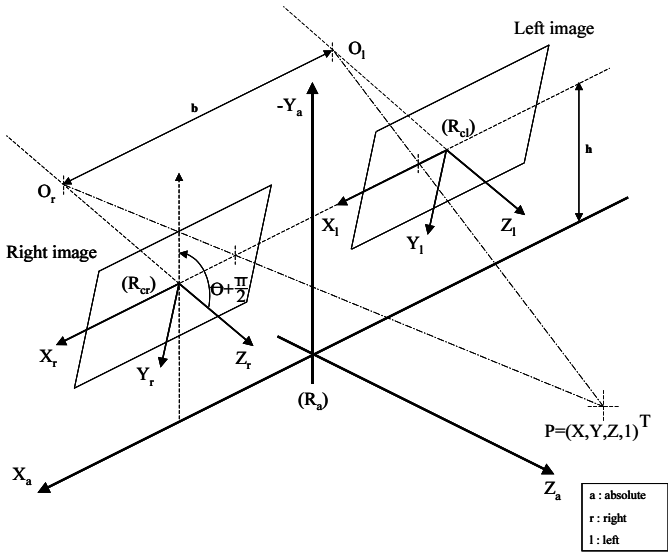


Fig. 1. The stereo sensor and the coordinate systems used.

The other parameters on the figure are as follows:

- θ : is the angle between the optical axis of the cameras and the horizontal,
- h : is the height of the cameras above the ground,
- b : is the distance between the cameras (i.e. the stereoscopic base).

In the camera coordinate system, the position of a point in the image plane is given by its coordinates (u, v) . The image coordinates of the projection of the optical center will be denoted by (u_0, v_0) , assumed to be at the center of the image. The intrinsic parameters of the camera are f (the focal length of the lens), t_u and t_v (the size of pixels in u and v). We also use $\alpha_u = f/t_u$ and $\alpha_v = f/t_v$. With the cameras in current use we can make the following approximation: $\alpha_u \approx \alpha_v = \alpha$.

Using the pin-hole camera model, a projection on the image plane of a point (X, Y, Z) in R_a is expressed by:

$$\begin{cases} u = \alpha_u \frac{X}{Z} + u_0 \\ v = \alpha_v \frac{Y}{Z} + v_0 \end{cases} \quad (1)$$

On the basis of Fig. 1, the transformation from the absolute coordinate system to the camera coordinate system is achieved by the combination of a vector translation $\vec{t} = -h\vec{Y} + \varepsilon_i \frac{b}{2}\vec{X}$ (with $\varepsilon_i = -1$ in R_{cl} or 1 in R_{cr}), and a rotation around \vec{X} by an angle of $-\theta$. Let T_i denote the translation matrix, R the rotation matrix and $D_i = RT_i$. In homogeneous coordinates, the different transformation matrices are therefore:

$$T_i = \begin{pmatrix} 1 & 0 & 0 & -\varepsilon_i \frac{b}{2} \\ 0 & 1 & 0 & h \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta & 0 \\ 0 & \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2)$$

where i is equal to r, l (right or left).

It is necessary to perform a perspective projection in order to express fully the coordinates of the points in the image plane coordinate system. The perspective projection matrix M_{proj} is expressed as follows:

$$M_{proj} = \begin{pmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (3)$$

Finally, we obtain the matrix of transformation T_{ri} from the absolute coordinate system R_a to the image coordinate system i (i is equal to l or r):

$$T_{ri} = M_{proj} D_i \quad (4)$$

If P is a point with homogeneous coordinates $(X, Y, Z, 1)^T$ in R_a , its homogeneous coordinates in the image coordinate system i are:

$$p = T_{ri} P = (x, y, z)^T \quad (5)$$

We can then compute the non-homogeneous (u, v) coordinates of P as:

$$\begin{cases} u = \frac{x}{z} \\ v = \frac{y}{z} \end{cases} \quad (6)$$

B. Modelling the longitudinal road profile

In what follows we will consider that the road is modelled as a succession of parts of oblique planes in R_a with respect to the plane of the stereoscopic sensor. This model will allow us to extract the longitudinal profile of the road as a piecewise linear curve (see Section IV.A). Our main assumption is only that the profile curvature is of constant sign.

III. "V-DISPARITY" IMAGE

A. Construction

We suppose that a disparity map I_Δ has been computed from the stereo image pair (see Fig. 6 Up Left). For example, this map is computed with respect to the epipolar geometry; the primitives used are horizontal local maxima of the gradient; matching is local and based on normalized correlation around the local maxima.

Let H be the function of the image variable I_Δ such that $H(I_\Delta) = I_{v\Delta}$. We call $I_{v\Delta}$ as the "v-disparity" image (see Fig. 6 Up Right). H accumulates the points with the same disparity that occur on a given image line i . For the image line i , the abscissa u_M of a point M in $I_{v\Delta}$ corresponds to the disparity Δ_M and its grey level i_M to the number of points with the same disparity Δ_M on the line i : $i_M = \sum_{P \in I_\Delta} \delta_{v_P, i} \delta_{\Delta_P, \Delta_M}$ where $\delta_{i,j}$ denotes the Kronecker delta.

Once I_Δ has been computed, $I_{v\Delta}$ is built by accumulating the pixels of same disparity in I_Δ along the \vec{v} axis (see Fig. 6 Up).

B. Semi-global matching

For a given image line, the grey level of a point in $I_{v\Delta}$ expresses the number of coherent points with disparity Δ_M in the same image line of the pair of stereo images. The construction of $I_{v\Delta}$ in facts amounts to a global analysis of an image line in I_{G_r} : we determine how this image line matches the same image line in I_{G_l} , for different horizontal offsets, that is to say for different disparities. In this manner, the horizontal order constraint is satisfied and the matching is semi-global (global for an image line).

C. The image of a plane in "v-disparity" image

Let P be a point with coordinates $(X, Y, Z, 1)^T$ in R_a . From system (6), the ordinate of the projection of this point on the left or right image is $v_l = v_r = v$:

$$v = \frac{[v_0 \sin \theta + \alpha \cos \theta](Y + h) + [v_0 \cos \theta - \alpha \sin \theta]Z}{(Y + h) \sin \theta + Z \cos \theta} \quad (7)$$

Moreover, the disparity Δ of the point P is:

$$\Delta = u_l - u_r = \frac{\alpha b}{(Y + h) \sin \theta + Z \cos \theta} \quad (8)$$

From (7) and (8), the plane with the equation $Z = aY + d$ in R_a is projected along the straight line of equation (9) in the "v-disparity" image:

$$\Delta_M = \frac{b}{ah - d}(v - v_0)(a \cos \theta + \sin \theta) + \frac{b}{ah - d}\alpha(a \sin \theta - \cos \theta) \quad (9)$$

Thus, a surface which is formed by a succession of parts of planes is therefore projected as a piecewise linear curve.

IV. APPLICATION

A. Robust determination of the longitudinal profile of the road

A.1 Case of a flat road geometry.

The plane of the road is projected in $I_{v\Delta}$ as a straight line with mean slope 0.70 (considering that mean values are $b = 1$ m, $h = 1.4$ m, $\theta = 8.5^\circ$). The longitudinal profile of the road is therefore a straight line in $I_{v\Delta}$. Robust detection of this straight line can be achieved by applying a robust 2D processing to $I_{v\Delta}$. In our application we use a Hough transform, the bounds of Hough space depending on the extreme values of h and θ that are tolerated.

The pitching and relative height of the stereo sensor are dynamically estimated by simply extracting the straight line of the road (see Fig. 2). Let c_r denote the slope and v_{or} the value of v for $\Delta = 0$ on the straight line of the road; we can obtain the following expression for θ and h :

$$\theta = \text{Arc tan}\left(\frac{v_0 - v_{or}}{\alpha}\right) \quad (10)$$

$$h = b \frac{\cos \theta}{c_r} \quad (11)$$

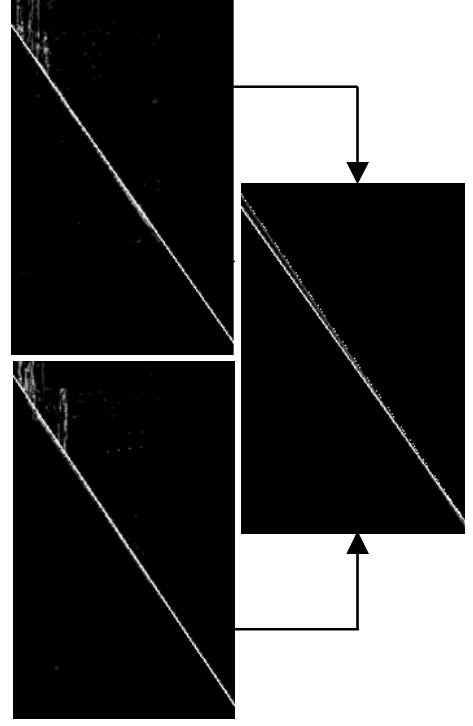


Fig. 2. Left Up and Down : extraction of the road profile from 2 "v-disparity" images of the same road sequence. Right : the two road profiles plotted on the same image show slope and position variations due to the vehicle suspensions.

A.2 Case of a non flat-earth road geometry.

The road is modelled as a succession of parts of planes. As a matter of fact, its projection in $I_{v\Delta}$ is a piecewise linear curve. Computing the longitudinal profile of the road is then a question of extracting a piecewise linear curve in $I_{v\Delta}$. Any robust 2D processing can be used. In our application we currently use the Hough Transform. θ and h can still be estimated. Indeed, in the vicinity of the vehicle, the road is planar. Thus, it is possible to extract the projection of the planar part of the road (which is a part of a straight line in the right partion of $I_{v\Delta}$) and estimate θ and h .

A.3 Extracting a piecewise linear curve using the Hough Transform.

The bounds of Hough space can be limited considering the maximal 3D road slope tolerated. In the considered area, the k highest Hough Transform values are retained (in our application, k is taken between 5 and 12). The k selected points correspond to k straight lines in $I_{v\Delta}$. The piecewise linear curve researched is either the upper (when approaching a downhill gradient) or the lower (when approaching an uphill gradient) envelope of the family of the k straight lines generated. We choose between these two as follows: $I_{v\Delta}$ is investigated along both curves extracted and a score is computed for each : for each pixel on the curve, the corresponding grey level in $I_{v\Delta}$ is accumulated. The curve is chosen with respect to the best score obtained (see Fig. 3).

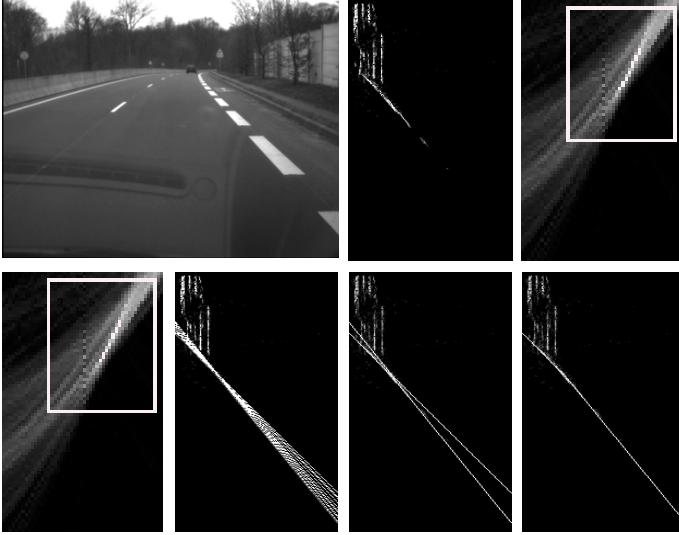


Fig. 3. Extracting the longitudinal profile of the road. (From Left to Right) Up : an image of the stereo pair corresponding to a non flat road geometry when approaching an uphill gradient, the corresponding $I_{v\Delta}$ image, the associated Hough Transform image (the white rectangle show the research area of the k highest values). Down : the set of the k straight lines generated, the computed envelopes, and the resulting longitudinal road profile extracted.

B. Computing the obstacle areas and the free space on the road

Once the longitudinal profile of the road has been extracted, the disparity values on the road surface are known for each scanning line (it is exactly the disparity value corresponding to the longitudinal profile curve extracted for the scanning line in question). Let Δ_i denote this value. Thus, it is straightforward to extract the obstacles from the disparity map I_Δ already computed: for a given scanning line, pixels whose disparity is equal to Δ_i are located on the road surface; other pixels belong to potential obstacles. On Fig. 4, note that the vehicle located at almost 80 meters is well perceived as a potential obstacle.

Once the obstacle areas have been computed, the free space on the road surface can be extracted (see also [6][7]), using a growing area algorithm (see Fig. 6 Down). The part of the image corresponding to the free space area can be used, for example, for extracting the lane-markings in more suitable conditions.

C. Dynamic estimation of the position of the line of the horizon

By definition, the points located on the horizon line are at a distance of infinity from the vehicle. Consequently, from (8), their disparity is nil. Furthermore, these points belong to the road. In "v-disparity" space, the position of the horizon line is therefore given by the value of v corresponding to $\Delta = 0$ on the straight line of the road.

On Fig. 5, the maximum difference between the position of the horizon line when manually estimated and when automatically estimated does not exceed 2 pixels.

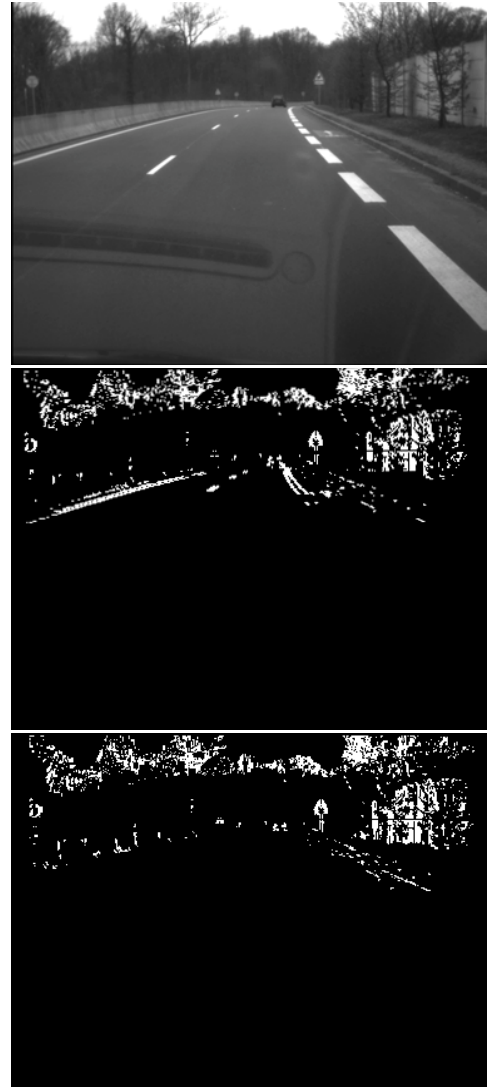


Fig. 4. Up: image of a non flat road geometry from the stereo pair. Middle: the corresponding obstacle areas computed when assuming that the road is planar. Down: the corresponding obstacle areas computed when there is no flat geometric assumption. In the first case, lane-markings located a few meters in front of the vehicle are considered to belong to obstacles while there is not such problem in the second case.

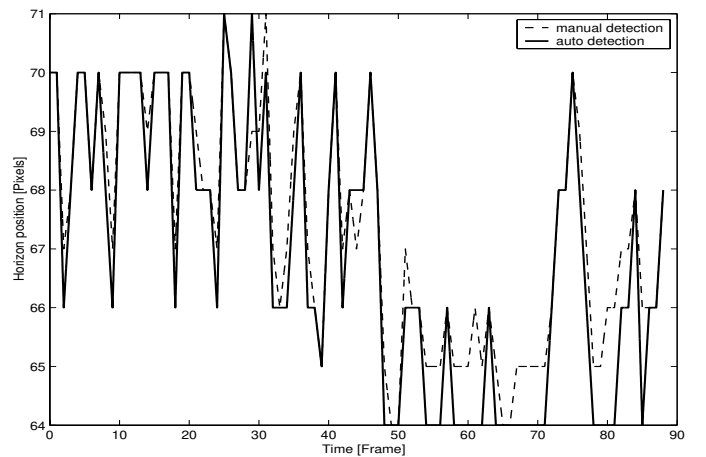


Fig. 5. Comparison between manual and automatic estimates of the variations in the position of the horizon over time

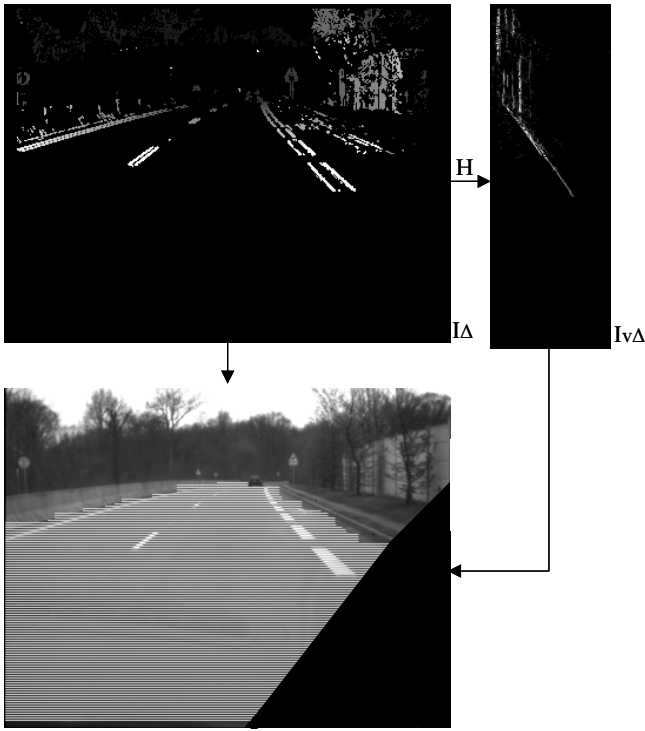


Fig. 6. Up Left: the disparity map (the grey level increases the higher the disparity). Up Right: the corresponding "v-disparity" image computed from the disparity map. Down: in grey, the free space on the road surface. The border of the black area on the Low Right corresponds to the longitudinal profile of the road extracted from the "v-disparity" image and is not a straight line but a piecewise linear curve.

D. Computing the position of tyre-road contact points

An obstacle is characterized by a vertical plane. With the mean camera position used, the corresponding straight line is quasi-vertical in $I_{v\Delta}$. It can be extracted by locating maxima on a histogram that sums the pixels in each column of $I_{v\Delta}$.

The tyre-road contact point of a vehicle is located at the intersection between the surface of the road and the vertical plane, and therefore, in $I_{v\Delta}$ it is the intersection between the longitudinal profile of the road and the quasi-vertical straight line corresponding to the rear of the vehicle (see Fig. 7). Fig. 8 shows the result of the estimation of the tyre-road contact point on a non flat road geometry over time. Vehicle is located at about 80 m in front of our vehicle. It should be noted that even if the disparity value of the lower parts of the vehicles is erroneous, the position of tyre-road contact point will still be estimated correctly as the point of intersection between the "obstacle" plane projection and of the road longitudinal profile.

Bounding boxes on vehicles can be built on the basis of the tyre-road contacts (see Fig. 9). Right and left sides are computed on the basis of the "u-disparity" image (u is the abscissa of a pixel in the (u,v) coordinate system) which provide a top view of the (u,v, Δ) space.

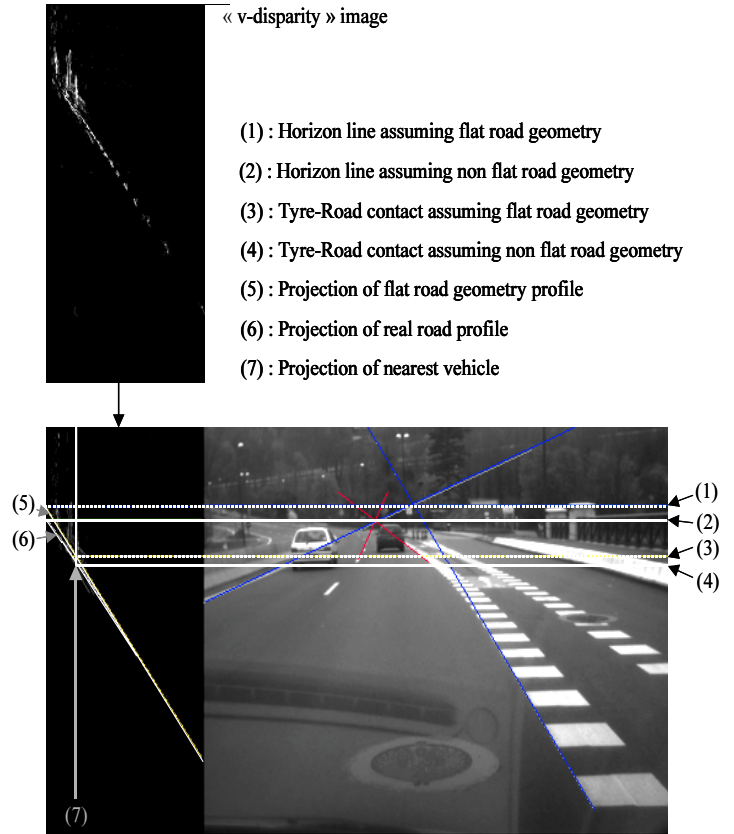


Fig. 7. The tyre-road contact of the nearest vehicle obtained from the "v-disparity" image.

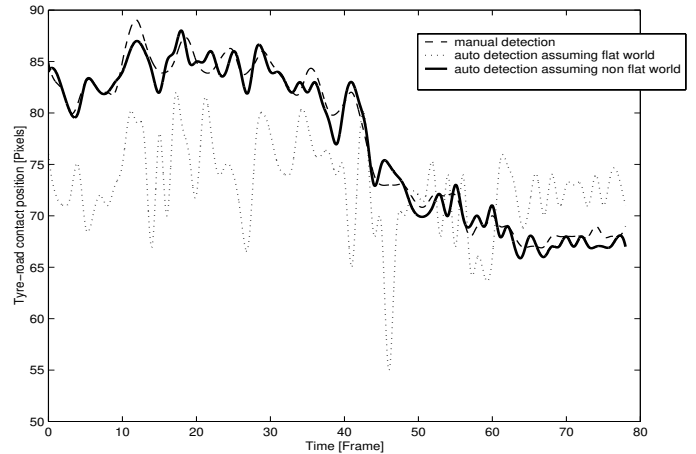


Fig. 8. Comparison between manual and automatic estimates of the variations in the position of the tyre-road contact point of the nearest vehicle over time.

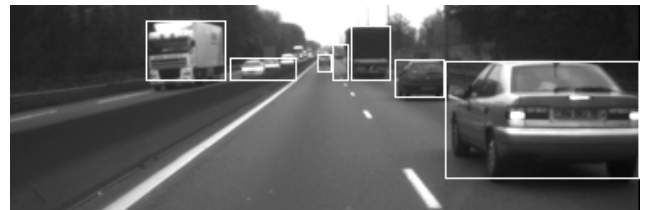


Fig. 9. A result of bounding boxes on multiple vehicles.

E. Some good properties of our method

The alignments in $I_{v\Delta}$ corresponding to the longitudinal profile of the road or to obstacles can be detected by robust methods such as the Hough transform. Thus, the presence of several alignments in $I_{v\Delta}$ merely affects the precision in which they are extracted. Furthermore, the identification of 2D alignments means that only the structures which are globally coherent in 3D space are selected and ensures that detection is robust to partial occlusions.

It should also be noted that:

- the construction of $I_{v\Delta}$ required no explicit extraction of coherent structures in the left and right images. This increases the robustness of our approach as the extraction of coherent structures in 2D images is often a source of errors. Furthermore, all the information in the disparity map I_Δ is exploited and the accumulation performed by H increases the density of the alignments in $I_{v\Delta}$. Any matching errors that occur when I_Δ is computed cause few problems as the probability that the points involved will generate coincidental alignments in $I_{v\Delta}$ is low.
- no back projection that is likely to reduce precision or amplify noise due to discretization is involved in the construction of $I_{v\Delta}$.
- $I_{v\Delta}$ corresponds to the image of a cross-section through the (u, v, Δ) space, in a vertical plane which is oriented in the direction of the optical axis of the camera. All the matched points are accumulated onto this plane. Thus, the road obstacles, which are orthogonal to this cross-section plane, generate dense alignments as they are observed edge on. The other objects -for example tree lines, buildings, etc.- only generate diffuse areas in $I_{v\Delta}$ as they are not orthogonal to the plane of the cross-section. For this reason the projection of these planes has little effect on the extraction of meaningful information (road profile, obstacles).
- our method works whatever the robust process used for computing the disparity map or for processing the "v-disparity" image.

F. Computational cost

The whole process for extracting the longitudinal profile of the road and computing the tyre-road contact points is performed within 40 ms with a current-day PC. The hardware used for the experiments is a Pentium IV 1.4 GHz running under Windows 2000. Images are grabbed using a Matrox Meteor II graphic card. The focal length of the lens is 8.5 mm. Image size is 380x289.

V. CONCLUSION

In this paper we have described a fast, accurate and robust method for detecting the road obstacles either on a flat road geometry road or on a non flat road geometry using stereo-vision. The detection process is based on the construction and the processing (through a Hough transform) of the "v-disparity" image. This image provides a 2D summary of all the information that is required in order to rapidly detect and robustly estimate the position

of obstacles even in the event of partial occlusion or errors committed during the matching process. Furthermore, it provides semi-global matching and reveals the matches which are the most coherent globally in the 3D road scene. The longitudinal profile of the road is extracted precisely. All other information concerning obstacles are then deduced in a straightforward manner : obstacle areas, free space on the road surface, position of tyre-contact points. This detection is performed without any explicit extraction of specific structures (road edges and lane-markings, etc.) but exploits all the relevant information in the stereo image pair. Computational time does not exceed 40 ms for images of size 380x289 on a current day PC with no special hardware.

The system will be tested on our experimental vehicles for Stop'n'Go applications. Future work will also be concerned with the accurate extraction of lane-markings, the detection of obstacles in our lane, and the construction of a 3D road model. The matching process could also be performed in a different manner: instead of basing this process on solely the maximum of correlation (which sometimes leads to the lost of good matches) it would be possible to consider the k best matches. Very fast disparity image computation techniques will also be tested, even at the cost of obtaining disparity maps of poorer quality, since our method is likely to cope with such poor disparity maps. The generalization of our method to any accumulation axis (and no longer the v axis) can lead to the estimation of roll and yaw rate values. Subsequent research will tackle this problem. The precise analysis of road scenes, including roll and yaw rate estimation could thus be achieved in real time.

ACKNOWLEDGMENTS

The authors would like to acknowledge Pr. Jean Devars's contribution (University of Paris VI). This work is partly funded by the European CARSENSE project and the french MICADO project.

REFERENCES

- [1] M.I.A. Lourakis and S.C. Orphanoudakis. Visual detection of obstacles assuming a locally planar ground. In *Proc. 3rd ACCV*, volume II, pages 527-534, 1998.
- [2] T. A. Williamson. A high-performance stereo vision system for obstacle detection. September 25, 1998. CMU-RI-TR-98-24. Robotics Institute Carnegie Mellon University. Pittsburg, PA 15123. Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
- [3] Z. Zhang, R. Weiss, and R.A. Hanson. Qualitative obstacle detection. In *Proc. of International Conference on Computer Vision and Pattern Recognition*, pages 554-559, 1994.
- [4] D. Koller, T. Luong, and J. Malik. Binocular stereopsis and lane marker flow for vehicle navigation : lateral and longitudinal control. Technical Report UCB/CSD 94-804, University of California at Berkeley, Computer Science Division, 1994.
- [5] H. Hattori and A. Maki. Stereo without depth search and metric calibration. Research & Development center, TOSHIBA Corporation. Kawasaki 212-8582, Japan. IEEE 2000.
- [6] M. Okutomi and S. Noguchi. Extraction of road region using stereo images. In *Proc. ICPR*, pages 853-856, 1998.
- [7] U. Franke and A. Joos. Real-Time stereo vision for urban traffic scene understanding. DaimlerChrysler AG. D-70546 Stuttgart. HPC : T728. IEEE 2000.