IN-VEHICLE OBSTACLES DETECTION AND CHARACTERIZATION BY STEREOVISION

Raphael Labayrade and Didier Aubert

LIVIC - INRETS/LCPC, 13 Route de la Miniere, Batiment 140,
78000 Versailles Satory, France
email raphael.labayrade@lcpc.fr, didier.aubert@inrets.fr
http://www.inrets.fr/js/ur/livic/rech_livic.htm

ABSTRACT

In this paper, we propose a fast and robust stereo algorithm to perform in-vehicle obstacles detection and characterization. The stereo algorithm used is called the "v-disparity"[1] algorithm which provides a suitable representation of the geometric content of the road scene. The stereo algorithm principle is described, and then the in-vehicle embedded system is presented. This system can be divided into two main stages. The first one deals with onboard road obstacles detection (the focus is put on obstacle areas and free road surface extraction), whereas the second one is about obstacles characterization (car/truck discrimination). For each stage, we first present the way in which the "v-disparity" algorithm is used, and then report representative experiments on real situations which show that our solution is accurate, reliable and efficient. In particular, both processes are fast, generic, robust to noise and bad conditions, and work even in the case of partial occlusions.

Keywords : Stereoscopic vision, In-vehicle obstacles detection, Obstacles characterization.

1. INTRODUCTION

In the context of Intelligent Vehicles Systems, in-vehicle road obstacles detection is an essential task. It must be performed in real time, robustly and accurately, without any false alarm and with a low detection failure rate. First, obstacles must be localized on the road; additional information such as obstacles characterization (car/truck discrimination, and also speed computation and trajectory recovery) can be interesting in order to predict their dynamic evolution.

Stereovision is often used to this purpose [1] [2] [3] [4] but algorithms are sometimes not precise or fast enough to be used efficiently. In this paper, we present a stereo algorithm that can be used as the basis of obstacles detection and characterization in the automotive context. The proposed algorithm provides robust 3D global data that can be used efficiently for in-vehicle real time applications where the stereo sensor is moving.

Our stereo algorithm is based on the construction and subsequent processing of the "v-disparity" image. This image is obtained from a pair of stereoscopic images, after the computation of a sparse and rough disparity map (by "rough", we mean that it can contain numerous false matches; on the other hand such a disparity map is computed very quickly). From the "v-disparity" image, it is then straightforward to detect specific planes or cylindrical surfaces in the scene in a robust, fast and effective way (even in the event of partial occlusion or false matches). Thus, a good representation of the 3D geometric content of the scene is obtained and these data are used as the input of the obstacles detection/characterization system.

The obstacles detection process can provide two different sets of data. The first one is a set of global obstacles in the road scene (cars, trucks). The second one is obstacle areas, which include these global obstacles, and also all the objects located above the road surface (small obstacles on the side of the road, etc.). Concerning the obstacles detection stage, the main novelty shown in this paper deals with the improvement of the obstacle areas thanks to the improvement of the rough disparity map, even when the road is paved or dirty.
The obstacles characterization process uses the set of global obstacles detected in the previous stage. Car/truck dicrimination is performed robustly.

The reminder of the paper is organized as follows. Section 2 details the models we have used with respect to the stereoscopic sensor and the domain of validity of our study. Section 3 deals with the construction of the "v-disparity" image from a pair of stereoscopic images and presents a robust method for detecting specific surfaces in the scene.

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

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

Section 4 presents the in-vehicle obstacles detection system and stresses the improvement of the disparity map (and subsequent obstacle areas) from the ”v-disparity” image. Section 5 is about the obstacles characterization system. The last two sections include representative experiment reports on real situations.

2. HYPOTHESES AND DOMAIN OF VALIDITY

2.1. Modeling of the stereo sensor

The two image planes are merely in the same plane and at the same height above the ground (see Fig. 1). This camera geometry means that rectified images are obtained directly (the epipolar lines are parallel).

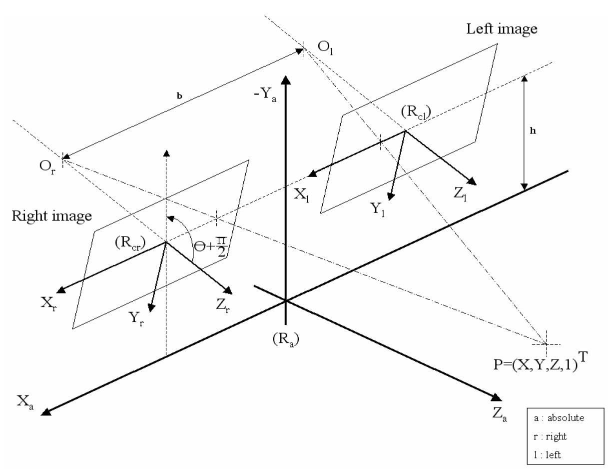

Fig. 1. The stereo sensor and the coordinate systems used

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). The other parameters on the diagram 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 coordinates in the image of the projection of the optical center will be denoted by (u_0, v_0), assumed to be at the center of the image. A projection on the image plane is expressed by:

$$\begin{cases} u = \frac{x}{z} \\ v = \frac{y}{z} \end{cases} \tag{1}$$

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 : $\alpha_u \approx \alpha_v = \alpha$.

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 $D_i = RT_i$ the result transformation matrix. In homogeneous coordinates, the different transformation matrices are therefore:

$$T_i = \begin{pmatrix} 1 & 0 & 0 & -\varepsilon_i \frac{b}{2} \\ 0 & 1 & 0 & h \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2}$$

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

$$D_i = \begin{pmatrix} 1 & 0 & 0 & -\varepsilon_i b/2 \\ 0 & \cos\theta & -\sin\theta & h\cos\theta \\ 0 & \sin\theta & \cos\theta & h\sin\theta \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4}$$

where i is equal to r, l (right and 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:

$$M_{proj} = \begin{pmatrix} \alpha_u & 0 & u_o & 0 \\ 0 & \alpha_v & v_o & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{5}$$

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):

$$T_{ri} = M_{proj} D_i = (t^t_{ijk})_{j=1..3,k=1..4} \tag{6}$$

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

$$p = T_{ri}P = (x, y, z)^T \tag{7}$$

On the basis of (1), we can then compute the (u, v) coordinates of P.

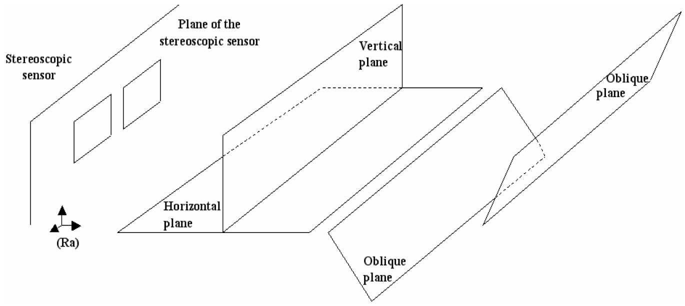

Fig. 2. The domain of validity of the study

2.2. Domain of validity

The aim of this study is to segment the environment into planes which are horizontal, vertical or oblique with respect to the plane of the stereoscopic sensor, as shown in Fig.2. The normal of the studied planes is therefore in a vertical plane which is oriented with respect to the optical axis of the cameras.

In a cross-section of the scene in the optical axis of the cameras, the projection of any of the planes in question is a straight line. In the rest of this paper we will build and use a specific image in which the detection of straight lines will be equivalent to the detection of planes in the scene. Indeed, we will represent the v coordinate of a pixel towards the disparity Δ (performing accumulation from the disparity map along scanning lines) and detect straight lines and curves in this 2-D $v - \Delta$ image (denoted by $I_{v\Delta}$). The next paragraph deals with the mathematical matching between global surfaces in the 3D scene and straight lines in this image.

2.3. The image of a plane in the ”v-disparity” image

This section describes the projections of the different planes that are considered in the ”v-disparity” image (see Fig. 2).

Let P be a point with coordinates $(X, Y, Z, 1)^T$ in R_a. System (1) allows us to compute the projection of this point in both images. As the epipolar lines are parallel and follow the image scanning lines, we again have $v_r = v_l = v$ where:

$$v = \frac{[v_0 \sin\theta + \alpha\cos\theta](Y + h) + [v_0 \cos\theta - \alpha \sin\theta]Z}{(Y + h)\sin\theta + Z \cos\theta} \quad (8)$$

We can also deduce from this the disparity Δ of the point P:

$$\Delta_P = u_l - u_r = \frac{\alpha b}{(Y + h)\sin\theta + Z\cos\theta} \quad (9)$$

From (9) and (8) the plane of the equation $Z = aY + d$ is projected along the straight line of equation (10) in ”v-disparity” image:

$$\Delta = \frac{b}{ah - d}(v - v_0)(a\cos\theta + \sin\theta) + \frac{b}{ah - d}\alpha(a\sin\theta - \cos\theta) \quad (10)$$

From (9) and (8) the plane of the equation $Y = c$ is projected along the straight line of equation (11) in ”v-disparity” image:

$$\Delta = \frac{b}{(c + h)}(v - v_0)\cos\theta + \frac{b}{(c + h)}\alpha\sin\theta \quad (11)$$

3. ”V-DISPARITY” IMAGE CONSTRUCTION AND 3D SURFACE EXTRACTION

We suppose that a disparity map I_Δ has been computed from the stereo image pair. 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. For example, in the experiments shown in the next sections, the threshold of the horizontal gradient is 8, and the size of the correlation window is 9x1 pixels. Thus, the matching process is very simple and fast.

Once I_Δ has been computed, the ”v-disparity” image $I_{v\Delta}$ is built by accumulating the pixels of same disparity in I_Δ along the $\vec{v}$ axis.

From (10), extracting straight lines or curves from the ”v-disparity” image leads to extract 3D global surfaces in the scene. Any robust 2D processing can be used to this purpose, like the hough transform. Details are given in [5]. In what follows we will suppose that the ”v-disparity” image is built and that global surfaces have been extracted.

4. OBSTACLES DETECTION SYSTEM

The global obstacles detection process is described in [5]. The extracted global surfaces correspond either to the road surface, or to obstacles. All needed information for performing generic obstacles detection is then deduced in a geometric way: vehicle pitch, obstacles-road contact points, distances computation (see Fig. 3). The accuracy of the algorithm is evaluated [6]. An extension of the algorithm is presented in [7] and leads to the estimation of roll and yaw angles of the vehicle.

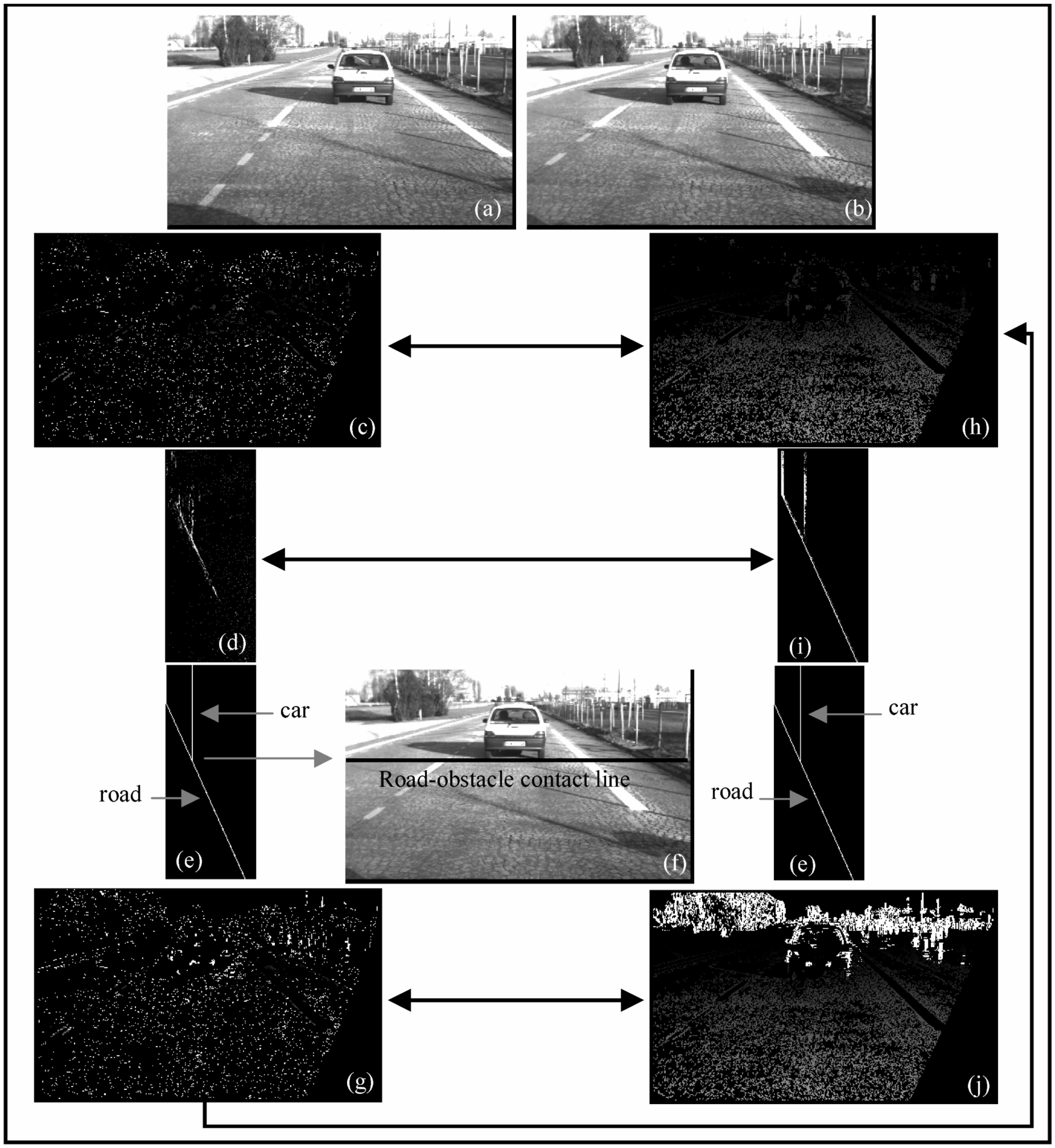

Fig. 3. In-vehicle obstacles detection framework including improvement of the disparity map and obstacle areas.
(a): left original image. (b): right original image. (c): rough disparity map computed from the images (a) and (b). (d): the grey level ”v-disparity” image corresponding to the disparity map (c). (e): extracted lines from the ”v-disparity” image (d). (f): road-obstacle contact line in the right original image (b). (g): obstacles areas (in white) computed from the disparity map (c) and the extracted lines (e). (h) improved disparity map computed from the images (a) and (b) and from the extracted lines (e). (i): the ”v-disparity” image corresponding to the improved disparity map (h). (j): obstacles areas (in white) computed from the improved disparity map (h) and the extracted lines (e). The grey pixels correspond to the disparity values on the road surface.
In the disparity maps (c) and (h), the disparity increases the higher is the grey level of the pixels. In (c), there is a lot of false matches but also good matches, especially on the road markings.
It is interesting to compare (c) and (h), (d) and (i), (g) and (j). (h), (i) and (j) are of far better quality.

4.1. Disparity map improvement and obstacle areas extraction

In order to fastly compute the "v-disparity" image, a sparse and rough disparity map has been built. This disparity map may contain numerous false matches, which is not a serious issue to extract global surfaces from the "v-disparity" image. However, in order to extract the obstacle areas, we should use a disparity map with very few false matches; otherwise these areas could be uncorrect. False matches are especially likely to occur in textured and repetitve areas, for instance on paved or dirty roads. So, in order to extract reliable obstacle areas on any road, false matches must be removed. This can be done thanks to the global surfaces extracted from the "v-disparity" image (see Fig. 3). Thus, we check wether a pixel of the disparity map belongs to any global surface extracted using the same matching process (cf. Section 3). If it is the case, the disparity value of the surface is mapped to the pixel. If not, that means that the pixel do not belong to any global surface: i.e. it belongs to a small obstacle. In practise, a short range of 2 pixels (along both u and v axis) around the disparity value of each global surface is investigated, in order to cope with some possible imprecision with respect to the extraction of the surface or with respect to the possible uncorrect sensor configuration and calibration mistakes.

Thus, the idea is to improve the disparity map from the geometric knoweldge about the scene obtained from the rough disparity map. In other words, the process is made of three stages (local, global, and local). The difference between the first and last local stages is that no knowledge is available in the first one, whereas the global geometric content of the scene is known in the last one. This incremental approach allows to obtain a disparity map of good quality and ensures computational speed. Such features could not have been achieved whitout the second global stage.

4.2. Results

We show the results from a paved road (see Fig. 3 [(h,j)] and Fig. 4), where the rough disparity map contain a lot of false matches. After improvement, the number of false matches on the road has been steadily reduced. Thus, the obstacle areas can be extracted in a far more reliable way, using a morphological operator to remove the very few remaining false matches, if needed. The disparity of some obstacle pixels can still be erronous, but that does not constitute any serious issue to extract the obstacle areas. From the obstacles areas, it is also possible to extract the free road surface in an accurate and efficient way, using a growing area algorithm: at the moment it is extracted as the convex area on the road limited by the obstacle areas. The improvement of the disparity map is performed within 100 ms. The whole obstacles detection process is performed within 150 ms using a 1.4 GHz processor. Image resolution is 380x288 pixels, and the focal lenght of the cameras is 8.5 mm.

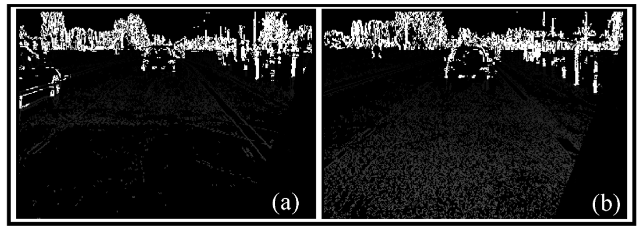

Fig. 4. Results of in-vehicle obstacles detection (obstacles areas in white). The grey pixels correspond to the disparity value on the road surface. (a): a car is overtaking on the left at the entrance of a paved area. (b): there is no noisy obstacle pixel in a repetitive textured road area.

5. OBSTACLES CHARACTERIZATION SYSTEM

Once the set of global obstacles is obtained, obstacles characterization is performed.

5.1. Car/truck discrimination

The discrimination is only based on the height of the vehicles: a vehicle is considered to be a car if its height is below 2 meters. On the contrary, a vehicle is considered to be a truck if its height is above 2 meters. In addition, we will consider that the height of a truck is below 5 m.

Thus, in the "v-disparity" image, we consider two different areas: the first one is located between the road surface and 2 meters above the road surface, the second one is located between 2 meters and 5 meters. These areas are determined as follows. The extraction of the profile of the road leads to an estimation of θ and h, using (11) with $c = 0$ (details are give in [5]). Thus, il is easy to deduce the equation of the projection of the planes $Y = -2$ and $Y = -5$ in the "v-disparity" image, using (11). The resulting areas are shown in Fig. 5 - (e), (f).

The question is then to know in which area are projected global obstacles already detected. To this purpose, for each global obstacle, we accumulate all the pixel grey values of the "v-disparity" image in each (car/truck) area. If the obtained value in the car area is above a threshold, it is then checked wether the obtained value in the truck area is above a threshold (See Fig. 5). Thus, the obstacle is characterised to be either a truck or a car. We do not use all the pixels of the disparity map, but only the ones located on the road: to this purpose, either a static mask or a dynamic area computed using a lane detector algorithm (inspired by [8]) is

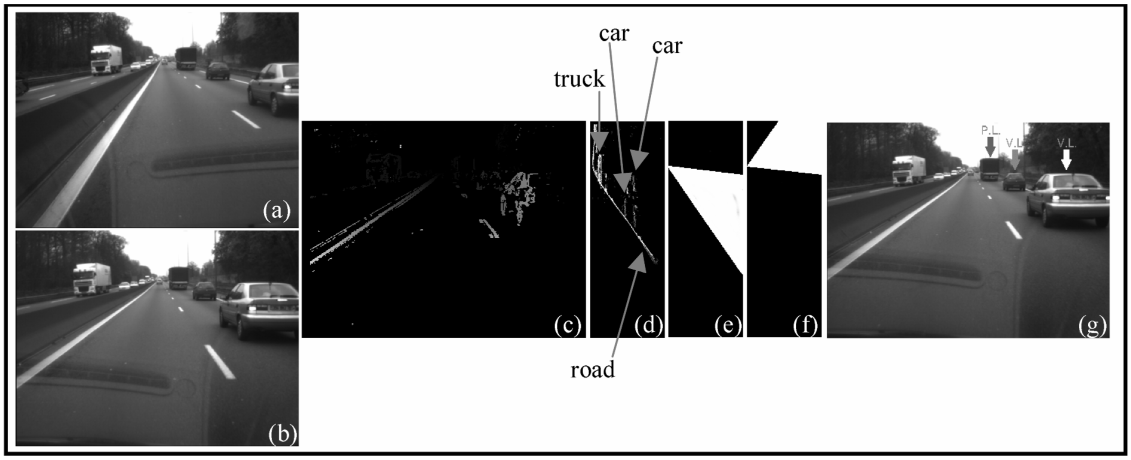

Fig. 5. In-vehicle obstacles characterization framework.
(a): left original image. (b): right original image. (c): disparity map. (d): ”v-disparity” image. (e): mask corresponding to cars (V.L.) area in ”v-disparity” image. (f): mask corresponding to trucks (P.L) area in ”v-disparity” image. (g): result of obstacles characterization.

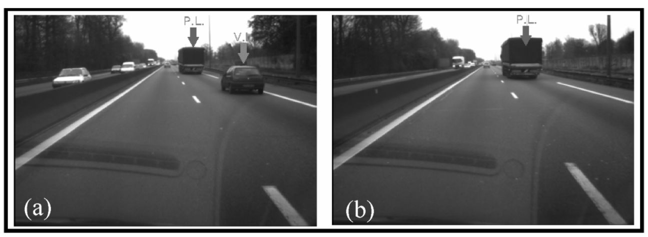

Fig. 7. Results of in-vehicle obstacles characterization: car(V.L)/truck (P.L.) (a): a car and an truck in the right lane. (b): the car has been overtaken.

used. In order to compute the speeds and recover trajectories of vehicles, a traking and associative algorithm can be used (cf. [9]).

5.2. Traffic monitoring system

It should be noted that the characterization algorithm can be used as a traffic monitoring system in the case where the stereo sensor is motionless and fixed above an highway, for instance. In this case, it is possible to perform vehicle counting, speed computation, accident detection. Experiments have been carried out. Representative results are shown in Fig. 6.

5.3. Results

In the tested sequences, the rate of good characterization is 100% (see Figs. 6 and 7). However, these results are not statistically valid since the amount of processed sequences is too low (785 pairs of images processed).

5.4. Some good properties of the system

The ”v-disparity” algorithm is characterised by good properties presented in [5] and [6]. First, since it is based on a geometric representation of the scene, the system works whatever can be the aspect of the road (shadows, dirty areas, oil, water, etc). Second, the longitudinal profile of the road is estimated so the system works whatever can be the road surface. Third, the system can detect any obstacle on the road since the detection is generic, even in the event of partial occlusion. Last, the global obstacle detection and characterization system is very robust to noise or bad conditions (weather, night) (experiments have been carried out in [6]) and runs in real time using no special hardware.

6. CONCLUSION

In this paper we have described a fast and robust stereovision algorithm (the ”v-disparity” algorithm) for determining the global geometric content of a scene. An in-vehicle application using these data has then been described. These application can be divided into two main stages. The first one is an in-vehicle obstacles detection system. An improvement of the obstacle areas has been proposed, so that the extraction of this area is reliable on any road, including paved roads. The second one is an in-vehicle obstacles characterization. Both stages profit from the good properties of the ”v-disparity” algorithm: computational speed, robustess towards noise and bad conditions, genericity. Experimental results show that the system works efficiently in the road context. Future work will be concerned with the quantitative evaluation of the new improvements towards a large amount of real situations.

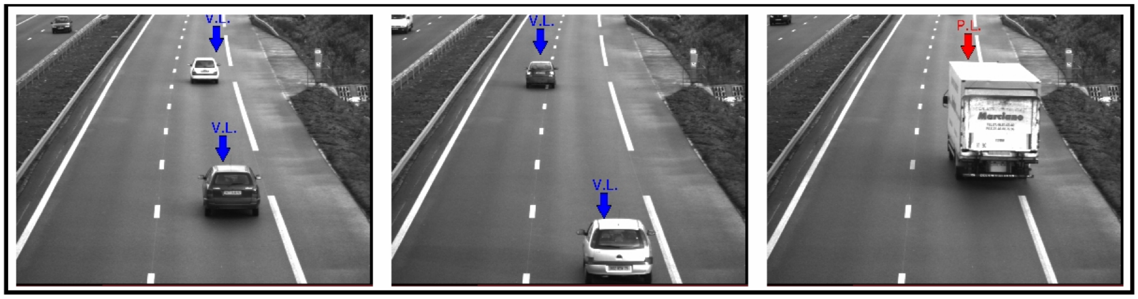

Fig. 6. Traffic monitoring results.

7. REFERENCES

[1] 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.

[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] U. Franke and A.Joos, ”Real-Time stereo vision for urban traffic scene understanding”, DaimlerChrysler AG. D-70546 Stuttgart. HPC: T728. IEEE 2000.

[4] H. Hattori and A. Maki, ”Stereo without depth search and metric calibration”, Research & Development center, TOSHIBA Corporation. Kawasaki 212-8582, Japan. IEEE 2000.

[5] R. Labayrade, D. Aubert, J. P. Tarel, ”Real Time Obstacle Detection on Non Flat Road Geometry through V-Disparity Representation”, IEEE Intelligent Vehicles Symposium, Versailles, June 2002.

[6] R. Labayrade and D. Aubert, ”Onboard Road Obstacles Detection in Night Condition Using Binocular CCD Cameras”, to be published in ESV 2003 Proceedings, Nagoya, Japan, 19-22 May 2003.

[7] R. Labayrade, D. Aubert, ”A Single Framework for Vehicle Roll, Pitch, Yaw Estimation and Obstacles Detection by Stereovision”, to be published in IEEE Intelligent Vehicles Symposium Proceedings, Columbus, June 2003.

[8] P. Coulombeau and C. Laurgeau, ”Vehicle Yaw, Pitch, Roll and 3D Road Shape Recovery by Vision”, IEEE Intelligent Vehicles Symposium, Versailles, June 2002.

[9] D. Gruyer, V. Berge-Cherfaoui, ”Multi-objects association in perception of dynamical situation”, UAI’99, Stockholm, Sude, 30 juillet-1 aot 1999.