Road Segmentation Supervised by an Extended V-Disparity Algorithm for Autonomous Navigation

Nicolas Soquet, Didier Aubert and Nicolas Hautiere

Abstract— **This paper presents an original approach of road segmentation supervised by stereovision. It deals with free space estimation by stereovision and road detection by color segmentation. The *v-disparity* algorithm is extended to provide a reliable and precise road profile on all types of roads. The free space is estimated by classifying the pixels of the disparity map. This classification is performed by using the road profile and the *u-disparity* image. Then a color segmentation is performed on the free space. Here is the supervision. Each stage of the algorithm is presented and experimental results are shown.**

I. Introduction

Autonomous navigation affords many useful applications: military unmanned vehicles, robotic exploration in extreme environments or driver assistances. It requires two major capabilities: the detection of potential obstacles and the estimation of the course of the road in the image.

Many systems have been designed to deal with obstacle detection in various environments. Radars [9] [18], laser range finder [10] [16], stereovision [19] [12] and multi-sensor fusion are used on structured roads. The 2005 DARPA Grand Challenge competition was successfully achieved on unstructured roads [6]. Most of the teams used lidar, radar or sonar [2] but no one succeed with a single vision-based system. Here will be an interesting and important evolution, because such systems have strong advantages such as low cost and passivity, which are key points for industrial or military matters. The TerraMax team completed the race with a stereovision-based obstacle detection [4] and the help of other active sensors [15]. This shows the potential use of stereovision. Thus we propose a single stereovision-based systems to detect the obstacles.

Road or road sides detection was widely explored too. Color segmentation [5] [14], texture analysis are commonly used [14]. Other systems use statistic model-based approach or morphological mathematics [1]. Many solutions were proposed for road or obstacle detection. However most of the time each system is specific to one single task.

This paper describes a new approach to deal with the perception of the environments for autonomous navigation. Our system finds a path in the free space that can be used by a vehicle. It performs obstacle and road detection. It is designed to work both on structured and unstructured roads by using a stereovision sensor. First it provides the free space using an extension of the *v-disparity* algorithm. Then a color segmentation extracts the area corresponding to the road. The segmentation is supervised by the stereovision because only the free space is segmented. The paper is organized as follows: section II presents the stereoscopic sensor and the set of equations used. Section III proposes adaptations of the *v-disparity* algorithm for the unstructured roads in order to obtain a robust road profile. Section IV explains how we estimate the free space. Section V shows how the free obstacle road course is extracted. Finally section VI presents experimental results obtained from real images.

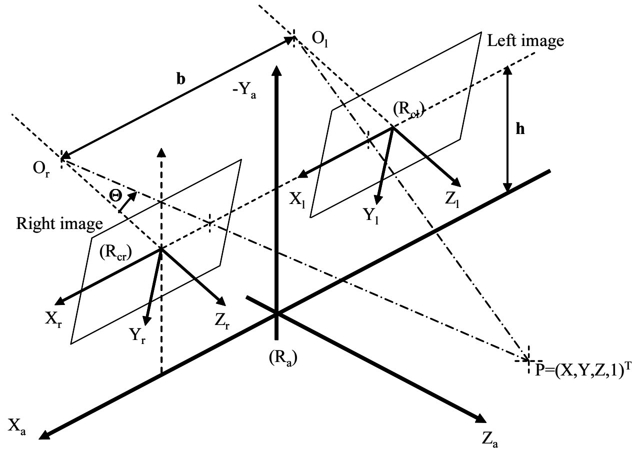

Fig. 1. Geometry of the stereoscopic sensor and coordinate systems (a: absolute, r: right, l:left).

II. Sensors and geometry

Our system uses a single stereovision sensor which is described in section II-A. The transformations between the disparity space (linked to the images) and the real space (linked to the vehicle) are described in the section II-B.

A. Stereovision sensor

The stereovision sensor (Fig. 1) uses two cameras positioned at the same height relative to the ground level. An adjustment is performed so that the epipolar lines are parallel and correspond to the scanning lines. The parameters described on the figure are:

- h: height of the cameras above the ground,
- b: stereo baseline (distance between the cameras),
- θ: pitch of the cameras (angle between the horizontal and the optical axis).

This work was sponsored by the French Defense Procurement Agency.

The authors are with LIVIC (INRETS / LCPC), Bât. 824, 14, route de la Minière, 78000 Versailles - France didier.aubert@inrets.fr nicolas.hautiere@lcpc.fr

The coordinates (u_p, v_p) give the position of a point P in an image plane. The intrinsic parameters of the cameras are the focal length of the lens (f) and the size of the pixels (t_u, t_v). According to the used cameras specifications, we can approximate: $t_u \approx t_v = t$. Then we use $\alpha = f/t$.
This sensor is used to estimate the free space, therefore a color camera is required for the road segmentation. In order to reduce the computation load, the color camera used is the right one of the stereoscopic sensor. It simplifies the computation because in that case the color segmentation and the stereovision algorithm are processed in the same coordinate system: the one linked to the right camera.

B. Coordinate systems

(R_a) is the absolute coordinate system. (R_r) is the coordinate system of the right image and (R_l) is the one of the left image. For a given point P, the R_a-coordinates are denoted by (X_p, Y_p, Z_p), the R_r-coordinates by (u_{rp}, v_{rp}) and the R_l-coordinates by (u_{lp}, v_{lp}). Δ_p is the disparity value and is equal to $u_{lp} - u_{rp}$. Thanks to the epipolar constraint we have $v_{rp} = v_{lp}$. We call disparity space the (u_{rp}, v_p, Δ_p) space.
Using the pin-hole camera model and the projection of the optical center (u_0, v_0), we obtain the following equations (1):

$$\left|\begin{array}{lcl} u_{lp} & = & u_0 + \frac{\alpha X_p + \alpha b/2}{(Y_p+h)\sin\theta + Z_p\cos\theta} \\ \\ u_p = u_{rp} & = & u_0 + \frac{\alpha X_p - \alpha b/2}{(Y_p+h)\sin\theta + Z_p\cos\theta} \\ \\ v_p = v_{rp} & = & \frac{[v_0\sin\theta + \alpha\cos\theta](Y_p+h) + [v_0\cos\theta - \alpha\sin\theta]Z_p}{(Y_p+h)\sin\theta + Z_p\cos\theta} \\ \\ \Delta_p & = & u_{lp} - u_{rp} = \frac{\alpha b}{(Y_p+h)\sin\theta + Z_p\cos\theta} \end{array}\right. \quad (1)$$

Those equations (1) describe the transformation between the image coordinate systems (R_r and R_l) and the real space (R_a).

C. Plane projection

The scene is often modelled by a set of planes in the real space. It is proven in [12] that a 3D plane estimation can be reduced to a 2D straight line estimation by using a well defined plane projection. Therefore if the scene is simply modelled with a planar road and vertical obstacles, then finding the good plane projection is sufficient to easily estimate the planes. It is shown in [12] that the *v-disparity* image is very powerful for such estimations.

1) The v-disparity image: is a plane projection. $\vec{u}$ is the normal and $(\vec{v}, \vec{\Delta})$ is the basis of the plane. The *v-disparity* image is computed by accumulating the pixels of same disparity along the $u-$axis. Let consider a pixel (denoted P) on the *v-disparity* image with coordinates (v_p, Δ_p). The intensity of P equals the number of pixels on the line v_p of the disparity map, having a disparity of Δ_p. Thanks to those accumulations the *v-disparity* image is very robust with regard to the noise. This image is used to extract the road plane and vertical obstacles (see [12]).

2) The u-disparity image: is a plane projection. $\vec{v}$ is the normal and $(\vec{u}, \vec{\Delta})$ is the basis of the plane. The *u-disparity* image is computed by accumulating the pixels of same disparity along the $v-$axis. Like the *v-disparity* image, the *u-disparity* image tolerates the noise thanks to the accumulation process. However it is not used to extract any plane. It is used to detect obstacle pixels. Indeed high intensity in the *u-disparity* image indicates that many pixels in the disparity map have the same disparity in a single column, thus they do not belong to the road. They correspond to vertical alignments in the real space and can be considered as obstacles, then vertical alignment detection in the *v-disparity* image is useless. Moreover using the *u-disparity* image is more efficient, because it allows to deal with some non vertical obstacles (like security rails along the road). [3] uses a similar approach for obstacle detection by marking as obstacles the regions with similar disparity.

III. Road profile

Our approach needs to estimate the free space. This one is extracted thanks to a robust road profile estimation. This estimation consists of two different stages. First the *v-disparity* algorithm provides a global road profile. Then classification and propagation phases give a precise road profile.

A. Global road profile estimation

The global road profile is estimated thanks to an improved *v-disparity* algorithm.

1) The basic algorithm: described in [12] enables the estimation of a road global profile. The algorithm was designed to work on structured roads. First the primitives are computed using horizontal local maxima gradient on both images and a single threshold. Then the disparity map is computed by local matching, based on normalized correlation around the local maxima with a left/right checking failure procedure [7]. The *v-disparity* image is computed and a global road profile is extracted using any straight line extracting procedure, such as the Hough transform (see Fig. 2). When we have tested this basic algorithm on unstructured roads, we have sometimes extracted wrong global road profile. It happened because the image primitives were badly distributed in the images. The Hough transform did not take the best alignment in the *v-disparity* image. An alignment of few points with high intensity were preferred to an alignment of a larger number of points with lower intensity. On Fig. 3, the global road profile extracted is wrong. In this example, the numerous pixels corresponding to the trees on the horizon line accumulate to form a vertical alignment in the *v-disparity* image. Its intensity is much higher than the alignment corresponding to the road profile. With structured roads we can find a threshold to select primitives which give good results. However it is harder to find such a threshold with unstructured roads. The

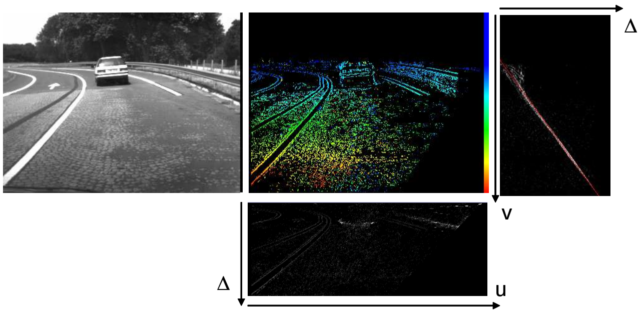

Fig. 2. *Top left:* an image of the stereo pair corresponding to a structured road. *Top middle*: the disparity map. *Top right:* the *v-disparity* image with the global road profile colored in red. *Bottom:* the corresponding *u-disparity* image. The images were computed with a gradient threshold of 6.

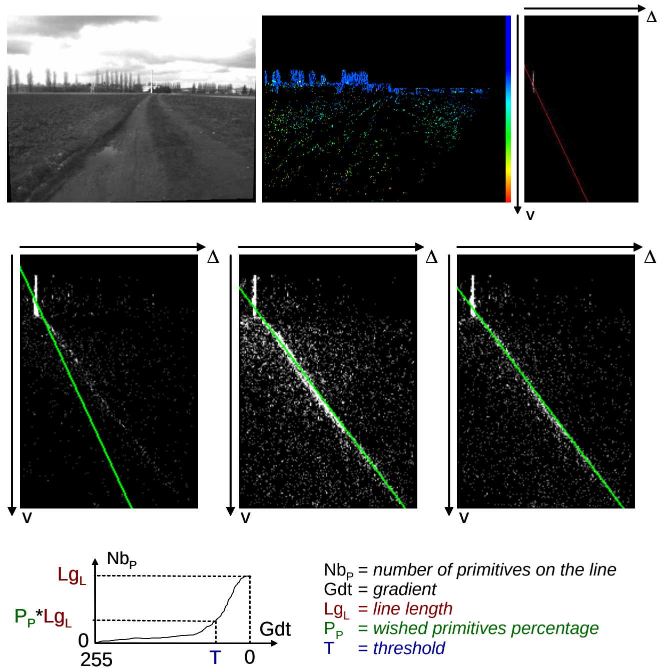

Fig. 3. *Top: left to right:* an image of the stereo pair corresponding to an unstructured road, the disparity map and the *v-disparity* image with the global road profile in red. The images were computed with a gradient threshold of 6. *Middle: left to right: v-disparity* image of a non structured road and the global profile in red for a threshold of 6, 2 and adaptive ones. *Bottom:* illustration of the cumulative histogram.

interval search of the plane parameters in the Hough space can be bounded to specific intervals, but it does not solve the problem in all cases because a bad distribution can generate a wrong estimate close to the good one.

2) Adaptive thresholds: are used to deal with this issue. Instead of setting a common threshold for both images, we set a percentage of primitives per line. This percentage is the same for all the lines of both images. Then we select a threshold for each line using a cumulative histogram (see Fig. 3). This algorithm is efficient to get a good primitives distribution. The Fig. 3 shows the results on the *v-disparity* and the global road profile. With a constant threshold of 6, the global road profile is wrong. With a constant one of 2, the profile is correct, but there is a lot of noise. With the adaptive thresholds, the profile is correct and the *v-disparity* image contains less noise. Most of the time the adaptive thresholds provide a correct global road profile. However in some cases a bad distribution still disturbs the Hough transform. Indeed the adaptive thresholds can not guarantee low intensity variations in the *v-disparity* image. It is the case, for instance, when an obstacle mask an important part of the image. In such cases, the Hough transform may not extract an alignment corresponding to the road profile, but a wrong one due to the high intensity of the obstacle pixels.

3) A column normalization: is processed to solve this problem. When an obstacle fits in a vertical alignment of the *v-disparity* image, we normalize each column of the *v-disparity* image. The intensity of each pixel is divided by the highest intensity in its own column. This way the influence of the pixels corresponding to obstacles is decreased and the longest alignment is more relevant. Since the longest alignment in the *v-disparity* image suits the road profile, the Hough transform is more reliable.

B. Precise road profile extraction

The enhanced algorithm provides an efficient way to extract the global road profile. Such a profile is not suitable for unstructured roads, because one straight line in the *v-disparity* image corresponds to a planar road. We need a better accuracy (roll, yaw, pitch or road slant) with unstructured environments. Some techniques were designed to solve this problem. The road can be modelled as a succession of planes. The roll, pitch and yaw can be estimated by an extension of the *v-disparity* algorithm to other planes extraction [11] or a local analysis of the *v-disparity* [13]. However our technique requires global road profile estimation, which can also be improved by looking for k distinct maxima in the Hough map for a piecewise linear estimate. Our approach to obtain a more precise profile of the road is original and different to the previously described ones [11] [13]. It exploits the *v-disparity* image to the maximum. The goal is to detect all the pixels belonging to the road on the *v-disparity* image. Thus we extract the exact shape of the road profile in the *v-disparity* image using a propagation algorithm. The algorithm consists of two different phases: initialization and propagation.

1) The initialization phase: consists in extracting the global profile as previously described. Then the pixels on the *v-disparity* image belonging to the global profile are selected. Only the pixels with high intensity value are kept. Those pixels are the basis for the next propagation phase. This method preserves the robustness of the *v-disparity* representation, because we keep the most cumulating pixels and we remove the noisy ones.

2) The propagation phase: needs two parameters: a threshold on the intensity of the *v-disparity* pixels to keep only the significant ones and remove the noise and a specific pattern to give the direction of the propagation. We start

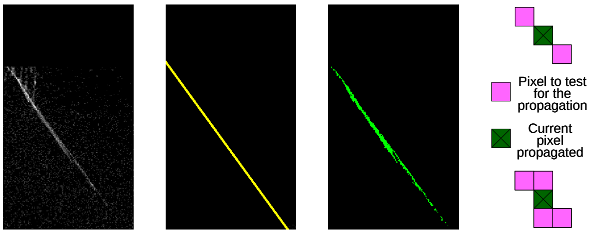

Fig. 4. *Left to right: v-disparity* image, the global profile in yellow color, the precise one in green color and two examples of possible patterns. *Concerning the patterns:* the bottom one is unsuitable because it allows propagation in the column of the current pixel. Indeed it may propagate an obstacle which must not belong to the precise profile. The top one is simple and appropriate because it allows a propagation in directions corresponding to the orientation of the profile.

the propagation from the seed pixels. For each tested pixel we check if the intensity value is high enough (above the specified threshold). If it is the case, we add this pixel (denoted P) to the profile and we test all the pixels given by the used pattern centered on P. Otherwise we do nothing, P is left and we restart from another seed. We go on until each seed has been propagated. The obtained profile has a pixel precision. It does not allow robust yaw, roll or pitch estimation but it allows to take into account their variation. Then it improves the classification of the pixels of the disparity map.

IV. Free space

The precise profile is used to classify the pixels: the ones belonging to the road surface and the others supposed to be obstacles or noise. To perform this classification the *u-disparity* image is used as well.

A. Pixel classification

The classification phase does not process all the pixels of the image, because the disparity value is required. Thus it processes only the pixel of the disparity map which have a non-zero disparity value. That means that only the primitives, computed with the adaptive thresholds, which are matched, will be classified. Therefore for such a pixel (denoted P) the non-zero disparity value gives a projection point in the *u-disparity* image (denoted P_u) and a projection point in the *v-disparity* image (denoted P_v). Then the classification algorithm works as follow: if the intensity value in the *u-disparity* image of P_u is high then P is classified as an obstacle pixel. Otherwise if P_v belongs to the precise profile, P is classified as a road surface pixel, otherwise P is not classified (it is supposed to be noise). This classification phase does not consider all pixels. However the ones classified are very reliable due to the robustness of the *v-disparity* and *u-disparity* images. That is why they are used to extend the classification to the unclassified ones with a propagation phase.

B. Propagation of the free obstacle area

The propagation phase is straightforward and based on the robustness of the classification. The reliable pixel are classified into road surface and obstacles. An influence area is computed for each of them by using a two dimensional gaussian surface. Thus all pixels receive a contribution from the classified pixels. The value of each contribution depends on the distance to the classified pixel and its class. If it is a free space pixel the value is positive. Otherwise it is negative. The longer is the distance, the lower is the absolute value. Then all pixels cumulate the value of the contribution of all classified pixels. For instance an unclassified pixel surrounded by free space pixels has a positive sum and is labelled as road surface. On the other hand an unclassified pixel surrounded by obstacle pixels has a negative sum and is labelled as obstacles. When the sum equals zero the pixel is classified arbitrary as road surface. In this manner all unclassified pixels are labelled. This phase is not as robust as the classification one, but it is a good way to show how the classification using a precise profile is efficient. In order to increase the robustness we plan to add a checking procedure of the final propagation.

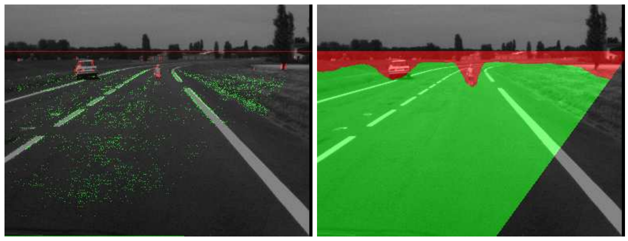

Fig. 5. *Left:* the result of the classification step, the green points correspond to the road classified pixels and the red ones to the obstacle classified pixels. *Right:* the result of the propagation one: the green area is the free space.

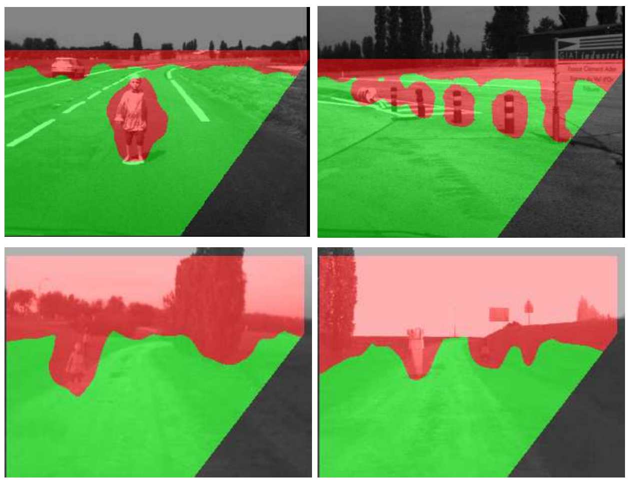

Fig. 6. Translucent green: the free space estimation, translucent red: the obstacle areas estimation.

V. Road segmentation

Once the free space is estimated, we have to find the road inside. We do not consider off-road environments but we deal with structured and unstructured roads. That means that a path must be distinguishable on the images. We use a color segmentation to extract the road.

A. *Work spaces*

Two different work spaces are considered:

- the image area where the segmentation is performed: it is the free space extracted by the stereovision algorithm. Here is the originality of the segmentation.
- the color space used: the hue and saturation coordinates of the HSL color space.

1) The free space: is the link between the stereovision and the segmentation. The color segmentation is processed on the right image of the stereoscopic sensor, however the whole image is not processed. Indeed the road can not be detected on the obstacles. Then only the part of the image corresponding to the surface of the road is processed. This area is given by the free space extracted by the stereovision algorithm (see IV). Thus the stereovision indicates where to search for the road. That is why we can say that the segmentation is supervised by the stereovision. Nevertheless the segmentation algorithm can work without the stereovision step. Then the free space corresponds to the entire image and the whole image is processed.

2) The color space: is essential for the segmentation. A two dimensional color space is used. We work in the HSL color space, but we keep only the hue (H) and saturation (S) coordinates for the segmentation. The luminance (L) coordinate is just used by the stereovision algorithm. Doing this allows us to deal with the shadows in a simple manner. Lying shadows on the road are useful for the stereovision algorithm, because we increase the numbers of primitives to match. However shadows and others light variations (reflection effects due to films of water) disturbed the color segmentation. Those light variations are mainly present on the luminance coordinate. We want to differentiate the road from the bank independently on the light conditions. The hue and saturation coordinates are less dependant and so enable a segmentation that is more robust to the light variations. Working on the hue and saturation coordinates does not completely remove the problems due to the light variation, but it is a simple way to be less disturbed.

B. *Color segmentation*

The segmentation is performed by the ISODATA clustering algorithm on the hue and saturation distribution (2D histogram) of the given image. The ISODATA algorithm is an iterative procedure working like the K-mean algorithm. First arbitrary initial cluster vectors are assigned. The second step classifies each point to the closest cluster. In the third step the new cluster mean vectors are calculated based on all the points in each cluster. Second and third steps are repeated until the difference between the iteration is small. This difference can be measured either by the distance mean cluster vectors have changed from one iteration to another or by the percentage of pixels that have changed between iterations. Compared to the K-mean algorithm, the ISODATA algorithm has some further refinements: splitting and merging of clusters (see [8]). Clusters are merged if the centers of two clusters are closer than a certain threshold. Clusters are split into two different clusters if the cluster standard deviation exceeds a predefined value. In short the ISODATA algorithm is similar to the k-mean algorithm excepted that the ISODATA algorithm allows for different number of clusters while the k-mean assumes that the number of clusters is known a priori.

C. *Road estimation*

1) The clustering: is straight forward. First we check for each pixel of the image to which cluster it belongs to. Then we have to decide which cluster correspond to the road and which one does not. This step is done by a simple classification algorithm. We define an isosceles triangle in the image. The base is lying on the bottom line of the image and the top is the middle of the horizon line. This line is determined by the stereovision algorithm. This triangle is a simple model of the road. We consider that the main part of the triangle belong to the road. Thus for each cluster we count the number of pixels inside the triangle and the number of pixels outside. If there are more points of the cluster outside, then the cluster is classified as non road, otherwise it is classified as road. Therefore all the clusters and pixels are classified. The clustering can be improved by using the results of the previous frame.

2) Some results: of the color segmentation and classification are shown on figure 7. On those images the segmentation is performed without stereovision. Thus we do not have the free space and the color segmentation is processed on the whole image.

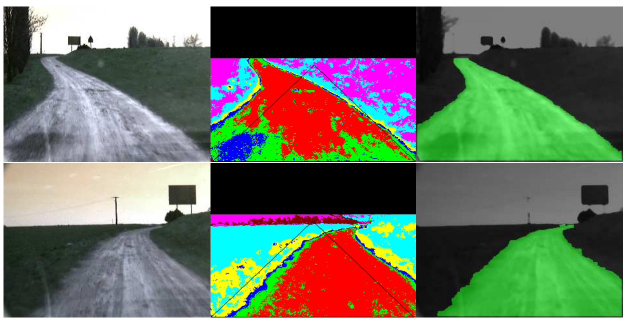

Fig. 7. *Left to right:* original image, clusters obtained, result of the classification.

VI. Results

Each step of our algorithm has been implemented and tested separately with real video sequences on a 2.4 GHz computer. We have already shown in the previous sections the results of the free space estimation (see IV) and the results of the color segmentation (seeV). The figure 8 shows some results obtained with the complete algorithm in various environments. In most cases the free space is well estimated. However even if the classification phase gives good results, the propagation phase can fail because of a lack of classified pixels. The number of primitives can be increased with the adaptive thresholds to obtain a denser disparity image. This would increase the computation load and prevent real time processing (30 fps). Naturally more textured are the images, better are the results. The color segmentation works as well. We obtain on the images a possible path without obstacles. Unfortunately our algorithm is sequential. The color segmentation is performed after the estimation of the free space. Thus the frame processing time increases again. Currently we can not stand real time processing and guarantee good results at the same time.

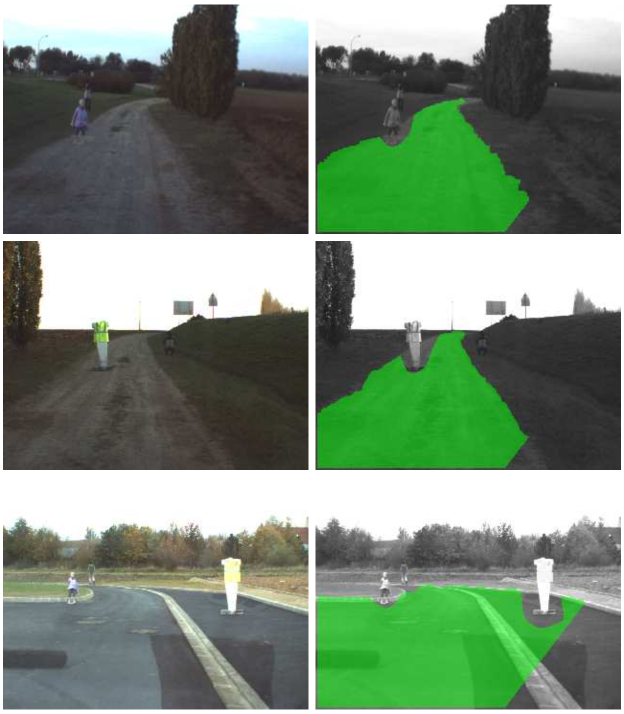

Fig. 8. *Left:* original images. *Right:* translucent green (or light translucent grey value): the road estimation.

VII. Conclusion

This paper presented an original method to perceive the environments for autonomous navigation. Our system detect the road in the free space thanks to a segmentation supervised by stereovision. It uses the robustness of the *u-disparity* and *v-disparity* representations. First the original *v-disparity* algorithm is extended to unstructured roads and enables the extraction of a precise road profile. This profile does not provide robust estimation of the roll, yaw or pitch, but it takes into account road variations (yaw, pitch, bumpy, damaged) and allows to work on any types of roads. The pixels of the disparity map are classified as road surface or obstacle. This phase gives very good results. The propagation of the free obstacle area labels all the remaining pixels and gives the free space. This last phase can be improved by using morphological processing (opening and closing operations) to remove small artefacts or to better emphasize the obstacles. The color segmentation is processed on the free space. Then the result of the stereovision is the input of the color segmentation. The supervision relies on this link. The segmentation is performed on the hue and saturation coordinates of the image by the ISODATA algorithm. An isosceles triangle is used to simply model the road and classify the cluster as road or not. The road segmentation can be easily improve. Indeed the use of the ISODATA algorithm affords the possibility to add other color or texture attributes because it is well designed for multi spectral segmentation. Vanishing point estimation [17] can be used as well. Other developments are considered. Working on the disparity map density, using the results of the previous frame and reducing the computation time are key improvements that we plan to carry out.

References

[1] R. Aufrere, R. Chapuis, V. Marion, and C. Lewandowski. Road sides recogintion in non-structured environments by vision. In *Intelligent Vehicle*, Parma, Italy, June 2004.

[2] R. Behringer, S. Sundareswaran, B. Gregory, R. Elsley, B. Addison, W. Guthmiller, R. Daily, and D. Bevly. The darpa grand challenge - development of an autonomous vehicle. In *Intelligent Vehicle*, Parma, Italy, June 2004.

[3] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri. Obstacle detection with stereo vision for off-road vehicle navigation. In *Machine Vision for Intelligent Vehicles*, San Diego, USA, June 2005.

[4] A. Broggi, C. Caraffi, P. P. Porta, and P. Zani. The single frame stereo vision for reliable obstacle detection used during the 2005 darpa grand challenge on terramax. In *Intelligent Transportation Systems*, Toronto, Canada, Sept. 2006.

[5] J. Crisman and C. Thorpe. Unscarf: A color vision system for the detection of unstructured roads. In *International Conference on Robotics and Automation*, Sacramento, USA, April 1991.

[6] DARPA. Grand challenge 2005. http://www.grandchallenge.org.

[7] G. Egnal and R. P. Wildes. Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on pattern analysis and machine intelligence*, 24(8):1127–1133, 2002.

[8] J. R. Jensen. Introductory digital image processing: A remote sensing perspective. In *Prentice Hall*, 1996.

[9] T. Kato, T. Tanizaki, T. Ishii, H. Tanaka, and Y. Takimoto. 76 ghz high perf. radar sensor featuring fi ne step scanning mechanism utilizing nrd technology. In *Intelligent Vehicles*, Tokyo, Japan, June 2001.

[10] A. Kirchner and A. Ameling. Integrated obstacle and road traking using a laser scanner. In *Intelligent Vehicles*, USA, Oct. 2000.

[11] R. Labayrade and D. Aubert. A single framework for vehicle roll, pitch, yaw estimation and obstacles detection by stereovision. In *Intelligent Vehicle*, Columbus, USA, June 2003.

[12] R. Labayrade, D. Aubert, and J. Tarel. Real time obstacle detection in stereovision on non flat road geometry through v-disparity representation. In *Intelligent Vehicle*, Versailles, France, June 2002.

[13] V. Lemonde and M. Devy. Obstacle detection with stereovision. In *Mechatronics and Robotics*, Aachen, Germany, Sept. 2004.

[14] D. Mateus, J. G. Avina, and M. Devy. Robot visual navigation in semi-structured outdoor environments. In *International Conference on Robotics and Automation*, Barcelona, Spain, April 2005.

[15] U. Ozguner, K. A. Redmill, and A. Broggi. Team terramax and the darpa grand challenge: A general overview. In *Intelligent Vehicle*, pages 232–237, Parma, Italy, June 2004.

[16] M. Parent and M. Crisostomo. Collision avoidance for automated urban vehicles. In *Intelligent Vehicles*, Tokyo, Japan, June 2001.

[17] C. Rasmussen. Texture-based vanishing point voting for road shape estimation. In *British Machine Vision*, London, UK, Sept. 2004.

[18] T. Uebo, T. Kitagawa, and T. Iritani. Short range radar utilizing standing wave of mircowave or millimeter wave. In *Intelligent Vehicles*, Tokyo, Japan, June 2001.

[19] T.A. Williamson. A high performance stereo vision system for obstacle detection. In *PhD thesis*, Carnegie Mellon University, 1998.