Obstacle Detection with Stereo Vision for Off-Road Vehicle Navigation

Alberto Broggi, Claudio Caraffi, Rean Isabella Fedriga, Paolo Grisleri

Dipartimento di Ingegneria dell'Informazione
Università di Parma, I-43100 Parma, Italy - http://vislab.ce.unipr.it
E-mail: {broggi, caraffi, fedriga, grisleri}@ce.unipr.it

Abstract

In this paper we present an artificial vision algorithm for real-time obstacle detection in unstructured environments. The images have been taken using a stereoscopical vision system. The system uses a new approach, of low computational load, to calculate a V-disparity image between left and right corresponding images, in order to estimate the cameras pitch oscillation caused by the vehicle movement. Then, the obstacles are localized by stereo matching and mapped in real world coordinates. Experimental results on sequences taken from a moving vehicle (which partecipated to the DARPA Grand Challenge 2004) in different unstructured scenarios are then presented, to demonstrate the validity of the approach.

1. Introduction

The research that made possible the realization of our algorithm is related to the project that led to our participation in the DARPA Grand Challenge competition which took place in the Mojave desert (USA) on March 13, 2004. When DARPA (Defense Advanced Research Projects Agency) promoted the Grand Challenge, its goal was to test the state of the art and incentivate research in completely autonomous vehicles design. The Grand Challenge 2004 was concerned with the construction and the equipping of vehicles of any kind to travel autonomously the 142 miles that separate Barstow (California) from Primm (Nevada) in a maximum time of 10 hours: the prize for the winners amounted to one million US dollars. The exact path, which consisted of 1000 waypoints, would have been divulged only 2 hours before the race began. Our group developed the vision system of Team Terramax. The vehicle was equipped also with other sensors, like differential GPS, ladars, sonars...

Although laser sensors provide refined and easy-to-use information about the surrounding area, they also present some intrinsic limitations to their functioning. In fact, scanning the real world with only one degree of freedom involves that thin obstacles, like poles or fencing, if too far away, cannot be localized because they occupy a scanning angle lower than the laser sensor resolution. Plus, in correspondence to ground slope variation and sensible vehicle pitch it often happens that the scansion intersecates with the terrain surface, misclassifying it as an obstacle.

On the other hand, a vision system provides a large amount of data, but extracting refined information sometimes may be complex. Furthermore, other difficulties arise from the integration of data coming from many different sensors. These facts resulted in the majority of the participants at the competition that had invested resources in the development of a vision system not being able to totally exploit this functionality. Therefore, in such a practical application it is necessary to know what the vision module is supposed to do and what it is not, in order to simplify the problem.

The aim of our vision module is to compute a real world representation that allows the path-planner to find a safe path. The computation must be in real-time; a long distance (50m) obstacle detection range is needed in order to anticipate decisions. Furthermore, the vision module should overcome other sensor lacks, for example detecting thin obstacles. On the other side, no obstacle classification is needed, and small obstacle width estimation errors have no consequences. Besides, when we drive, we do not need to know the environment with millimetric precision.

The outline of this paper is as follows. Section 2 describes related work in pitch and obstacle detection. Our pitch and obstacles detection system is discussed in detail in section 3. Section 4 presents the experiments performed to demonstrate the feasibility and effectiveness of the approach. Section 5 gives the conclusions and future direction of this work.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

2. First considerations and related works

To obtain a real world representation from an image pair, it is necessary to know the cameras placement (see fig. 1) at the time of acquisition.

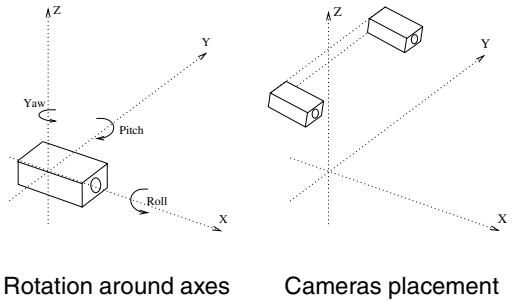


Figure 1. Real world frame of reference. The cameras are mounted on a rigid bar, and share the same pitch, yaw and roll angle.

In specific conditions, like on almost still vehicle or on highways, the static calibration angles are preserved at the time of acquisition. This is not true for a vehicle moving in extreme environments because of the bumps caused by the ground unevenness. In particular, the pitch oscillation causes a vertical discrepancy of the acquired images respect to the expected (on the basis of the static calibration) results.

Different approaches are possible to estimate the cameras angles at the time of acquisiton. For example, in [1] a yaw, roll and pitch estimation is made, referring to known characteristics of the environment (as lane width). Of course, this approach is not suitable for unstructured environments. Many works use a time correlation approach, tracking images recognized features and obtaining a visual odometry. In this field, [6] presents one of the most recent studies: it consists in a mono-camera method applied to both stereo images. On the contrary, we use a method known that uses a V-disparity representation, introduced by R. Labayrade and D. Aubert [4,5]: this method allows to obtain the cameras pitch angle at the time of acquisition from a single pair of stereo images.

A wide baseline is necessary to accomplish the goal of detecting far away obstacles. Therefore, most of the published studies about planetary exploration [2, 8] are not suitable for our applications. The usual approach in extremely unstructured terrain is to build a digital elevation map of the real world [11]. This method, due to its computational load, fails to fulfill the real-time requirements. Thus, we chose to compute a fast DSI using an area-correlation stereo method. We are not concerned by the "foreground fattening" effect [7], since it only cause the obstacles to appear slightly bigger.

3. Basic principles and implementation of the algorithm

In this section we discuss the process that starting from images acquisition (that includes the removal of the distorsion due to optical lenses[1]) leads to populate the real world map through the following elaboration steps:

1. V-disparity image computation;
2. pitch estimation;
3. disparity space image computation (DSI);
4. obstacles localization;
5. real world coordinates mapping.

3.1. V-disparity image computation in unstructured environment

The first step that leads to estimate cameras pitch is based on an approach similar to the one introduced by Labayrade [5], and consists in calculating the values of similarity (by a correlation measure), for different offset values (disparities), for each pair (left and right) of rows of the stereo images[2] at the same height (v coordinate). This operation enables us to produce a 3D graphic structured as it can be seen in fig. 2: the abscissa axis (d) plots the offset for which the correlation has been computed; the ordinates axis (v) plots the image row number; the intensity value is used as a third dimension, and settled proportional to the measured correlation, obtaining an image called V-disparity image (or correlation image).

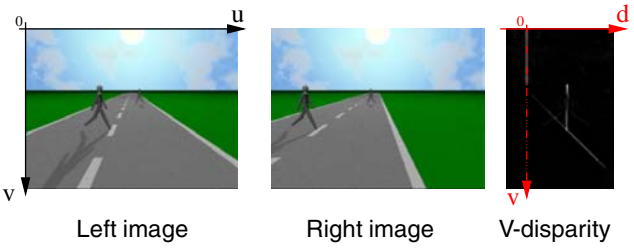


Figure 2. Images and V-disparity image frame of reference

Each planar surface in the field of view is mapped in a segment in the V-disparity image [5]. Vertical surfaces in

[1] Such feature is provided, together with the acquisition software and the real-world to image coordinates transformations, by the framework we used to develop the algorithm.

[2] Apart from a slight vergence, the cameras are in standard form [3], so the rows correspond to the epipolar lines.

the 3D world are mapped into vertical segments, while orizontal surfaces in the 3D world are mapped into slanted segments. E.g., the vertical segment in the center of the V-disparity image of fig. 2 is caused by the visible surface of the closest pedestrian. This surface is almost completely at the same distance from the cameras, so the corresponding segment has constant disparity. On the other hand, the slanted segment in the V-disparity image is caused by the linearly changing maximum correlation disparity among ground components. This segment, called "ground correlation line" in this study, contains the information about the cameras pitch angle at the time of acquisition (mixed with the terrain slope information). The first goal is then to obtain a V-disparity Image that allows to extract the ground correlation line.

First of all it is necessary to devise a feature in the images that allows to compute the correlation. Vertical edges allow to cope with the bias difference between the cameras and permit to extract the texture of the ground. For example, in the case shown in fig. 3, the Sobel filtered images (fig. 4) highlight two edges derived from the line on the right hand side; they are labeled with a and b.

Left image Right image

Figure 3. Stereo images of an asphalt road

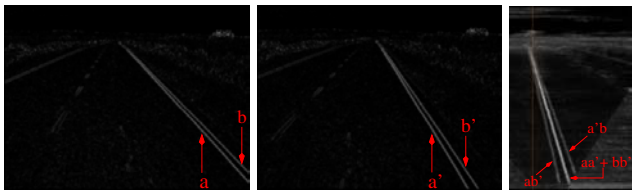


Figure 4. Edges images and their correlation image

In the correlation image, besides the ground correlation line, (coming from the contribute of the correlation between line a and a' and between line b and b') two other lines appear (ab' and $a'b$). They are generated by the high correlation between a and b' and between a' and b; of course this effect is undesirable because it has been originated by the matching of two objects that are not the same object.

To overcome this ambiguity, in this work we propose to consider, besides the absolute value of edges, also their phase: as can be seen in fig. 5, now a lines cannot match with b lines because of their different phases, and the correlation image is sensibly improved.

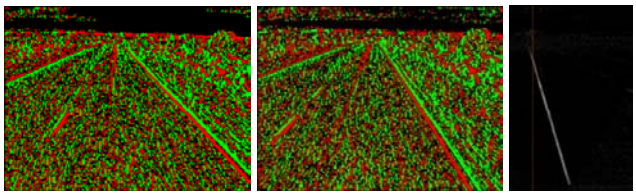


Figure 5. Correlation image computed from edges images taking into consideration the phase. Red represents negative phase of edges, green positive, black no edges.

Additional benefits can be obtained considering that in this step of the algorithm (ground correlation line highlighting) the information given by obstacles in the correlation image is considered as noise (see fig. 6b).

By ignoring the absolute value of edges and considering only their phase, the quantity of information carried by obstacles (tipically obstacles present very strong edges) is attenuated, while the weak edges produced by ground texture, that takes up the largest part of the image, yield to the greatest contribute in the V-disparity image. Theoretically every pixel, if having concordant phase with the pixel which it is compared to, will contribute with a positive vote, independently of the edge absolute value; otherwise, for discordant phase, a negative vote is produced. The sum of accumulated votes is normalized, producing the correlation value corresponding to a considered offset, following this formula:

$$corr(d) = \frac{\left(N_{match}(d)\right)^2}{N_L \cdot N_R} \qquad (1)$$

where:

- d is the disparity value used to compare the two rows;
- $N_{match}(d)$ is the number of phase matching between the left and right rows compared at disparity d;
- N_L and N_R are the number of non-black pixel respectevely in left and right rows.

A side effect of this operation is to lower the correlation computational cost (it is almost halved). In this study, this type of correlation is called "ternarized", since it comes from images mapped into a ternary domain (-1,0,+1). A further important remark is that this kind of correlation does not need any threshold.

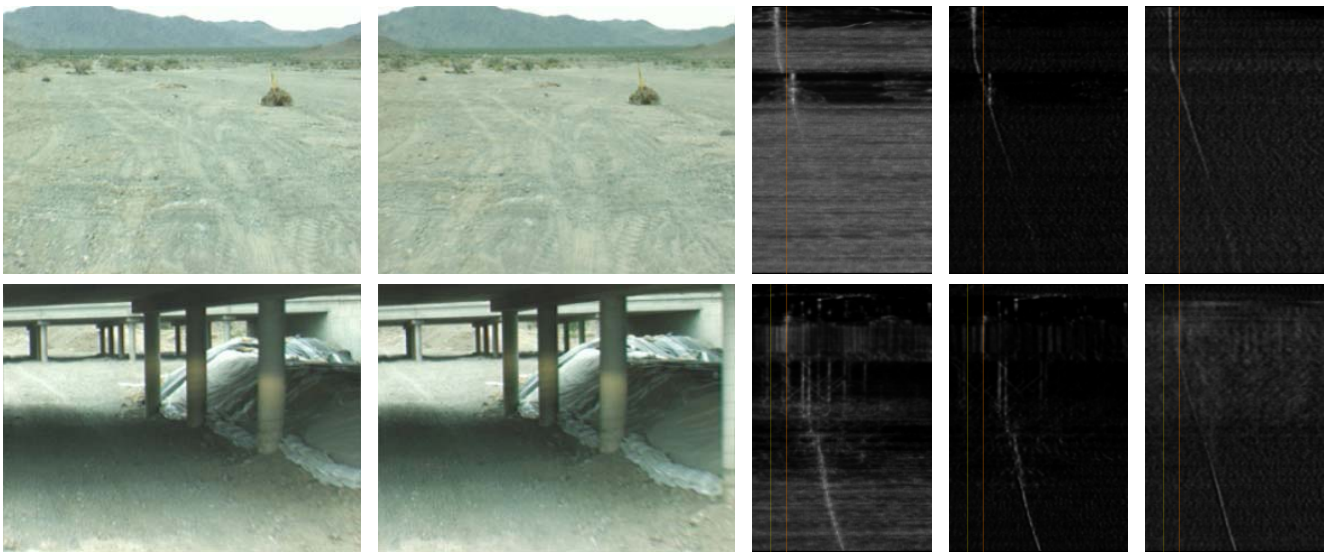


Figure 6. Correlation images in presence of obstacles: *(a)* **directly from edges image,** *(b)* **taking the phase into consideration,** *(c)* **ternarized.**

3.2. Pitch estimation

Using static calibration data, it is possible to compute the ground correlation line expected to appear in the correlation image when the vehicle is standing on a flat surface. It is also possible to compute the ground correlation line expected in case of different pitch angles, creating a set of candidate lines (see fig. 7). A constant ground slope approximation is made.

The behavior of the ground correlation line during a pitch variation is to oscillate, parallel to itself. Experimentally, we found out that the cameras height variation due to oscillation has neglectable effects. Accumulating the V-disparity image values along each of the candidate lines, it is possible to estimate the ground correlation line (choosing the line that accumulated the greatest value), and then the pitch of the cameras at the time of acquisition.

Furthermore, this method allows to compute the correlation image values only in correspondence to pixels belonging to the candidate lines, reducing the computational load. The whole correlation image, as seen in fig. 8, is computed only for debugging and displaying reasons.

The dynamic pitch information is used to determine a region of the images where to perform the search for obstacles: in fact the specific choice of the stereo baseline (wide enough to detect far away obstacles) does not allow to search in the immediate surroundings of the vehicle. We named this part of the images "region of interest". We assume that close obstacles will already have been localized then they were far away, or be seen by other sensors.

3.3. Disparity space image

The considered scenarios included obstacles like poles, underpass columns, bushes, trees, walls, artificial barriers, traffic signs, people, and other vehicles. The only feature they have in common is to come out from the ground plane and to have a vertical visible surface.

Labayrade [4,5] deals with obstacle detection by localizing vertical segments in the correlation image (for example, the correlation images of fig. 6b). For these disparity values, his algorithm studies in depth the images correlation, identifying the corresponding obstacle.

The problem of this approach is that it examines a limited number of disparities: it does not take into consideration that medium and small sized obstacles, occupying local image regions, may be not sufficiently visible in the V-disparity image, since it is built using global information. In fact, in unstructured environments the correlation image is suited to correctly detect only the ground plane, that occupies the largest portion of the image, but not to find obstacles, that are often only local features.

It is then necessary to think to a local approach, consisting of the search, for each point in the right image (chosen as reference image and restricted to the interest region), of its corresponding point in the left image. This leads to the creation of a disparity space image (DSI), where a disparity value is assigned to each region of the image depending on the most similar region that is found on the same row of the other image.

The disparity search range is bounded by using the (static

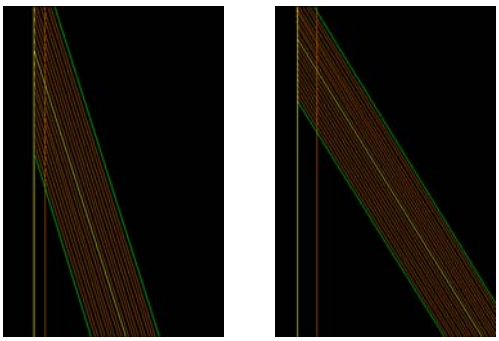


Figure 7. Lines generated by two different stereo baselines. The yellow slanted line corresponds to the ground correlation line obtained using the static calibration data, while the orange slanted lines are the ones expected varying the pitch value. The red vertical line indicates the 0 disparity value, the yellow vertical line indicates the disparity value of points at infinite distance; they do not overlap due to a slight convergence of cameras optical axes.

and dynamic) calibration data, making it dependent on the image v coordinate, in order to reduce the risk of wrong matches and lower the computational cost.

Assuming that small regions of each image have a similar homogeneus disparity, and considering that a similarity measure on a single pixel (which is just a value between -128 and 127) would lead to many false matches, we decided to split the region of interest of the images in small windows. We chose 3x3 squares in order to still be able to locate thin obstacles like poles. The similarity measure between windows is performed by the correlation formula shown in the following lines[3]:

$$Prod = \sum_{i=1}^{n}\sum_{j=1}^{m} LSquare[i,j] \cdot RSquare[i,j] \quad (2)$$

$$LQuad = \sum_{i=1}^{n}\sum_{j=1}^{m} \left(LSquare[i,j]\right)^2 \quad (3)$$

$$RQuad = \sum_{i=1}^{n}\sum_{j=1}^{m} \left(RSquare[i,j]\right)^2 \quad (4)$$

$$corr = \frac{Prod}{max(LQuad, RQuad)} \quad (5)$$

where:

[3]For a complete discussion of matching methods and correlation measurements, see [3, 7].

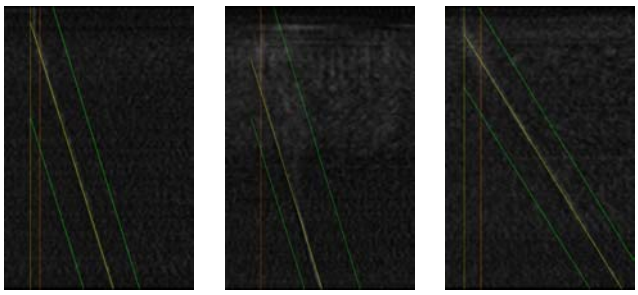


Figure 8. Examples of ground correlation lines individuation. The green lines bound the candidate lines set.

- $LSquare[i,j]$ and $RSquare[i,j]$ are the pixels at row i and column j of the examined windows in the left and right image respectively;
- n and m indicate height (number of rows) and width (number of columns) of the window in use (in this case $n = m = 3$).

3.4. Obstacle localization

The obstacles localization phase is performed by means of an aggregation step on the DSI: sufficiently wide regions at similar disparity are marked as obstacles. During this stage, a fine tuning of a threshold can provide good results. Anyway, no quantitative performance measurement in term of false positives and negatives has been computed yet, because of the difficulty of getting good ground truth test sets in unstructured environments.

The aggregation step is designed to ease pole localization, giving importance to the predominant disparity value of each image column. Fig. 9 shows an example of DSI computation and disparity aggregation that lead to obstacle localization. Fig. 10 shows successful localizations of obstacles in different scenarios. Different colors represent different distances, following the color coding in fig. 9.

3.5. Real world coordinates mapping

Finally, using the dynamic calibration obtained by the pitch estimation and the static calibration, it is possible to map the obstacles found in the images in a real world coordinates map. The correctness of this step is tightly dependent on the precision of static calibration data: in fact, during the other stages, even a weak calibration [12] permits anyway the algorithm to work: it is just essential that the candidate lines are generated parallel to the right static ground correlation line. On the contrary, during the final

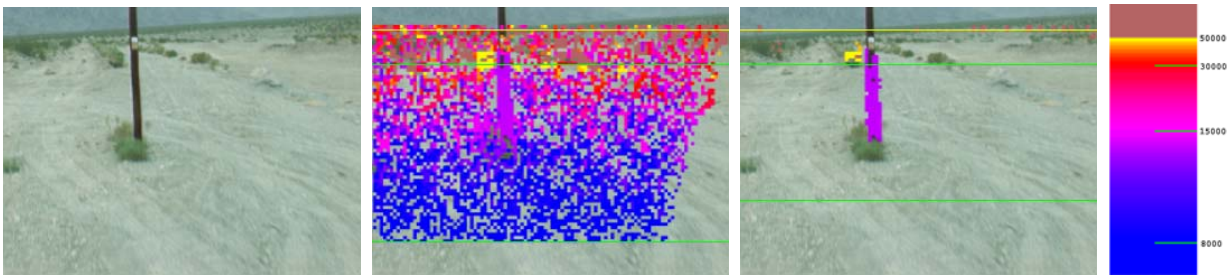


Figure 9. Example of DSI and obstacle localization. Different colors indicate different disparity values and, therefore, different distances from the cameras. In the color coding, the distances are expressed in millimeters. Note as the pole is characterized by a constant disparity in the DSI, and therefore is marked as an obstacle.

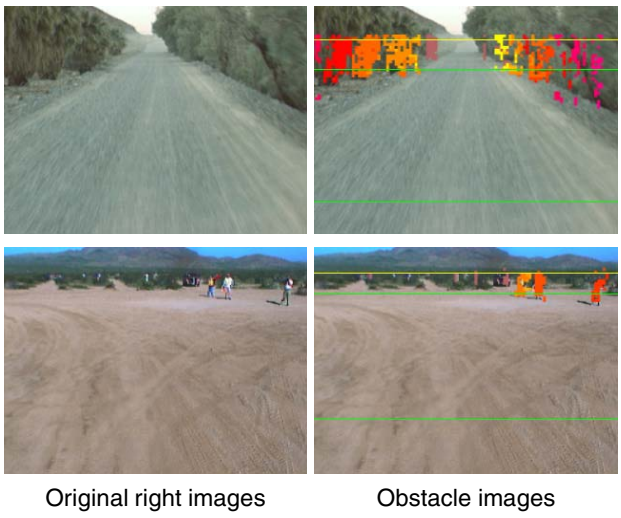


Figure 10. Examples of obstacles (trees and people) localization

real world mapping step, an accurate metric calibration [12] is mandatory: for example, a wrong relative (among cameras) yaw measurement can lead to sensible distance estimation error.

4. Results

The developed algorithm proved to produce good results in the conditions in which other sensors - such as lasers - failed: it was able to detect generic obstacles like isolated posts, thin fencing poles, trees and people. Some examples are shown in fig. 11 and in the submitted videos.

On a medium-hand processing system (a laptop with a P4-M 2,2 Ghz processor and 512 Mb of RAM) we obtained a total computational time of 64ms on 320x240 pixels images. This allows a real time application of the algorithm and, furthermore, permits to devise some enhancements to improve the results, for example introducing a better aggregation module to be used during obstacle localization.

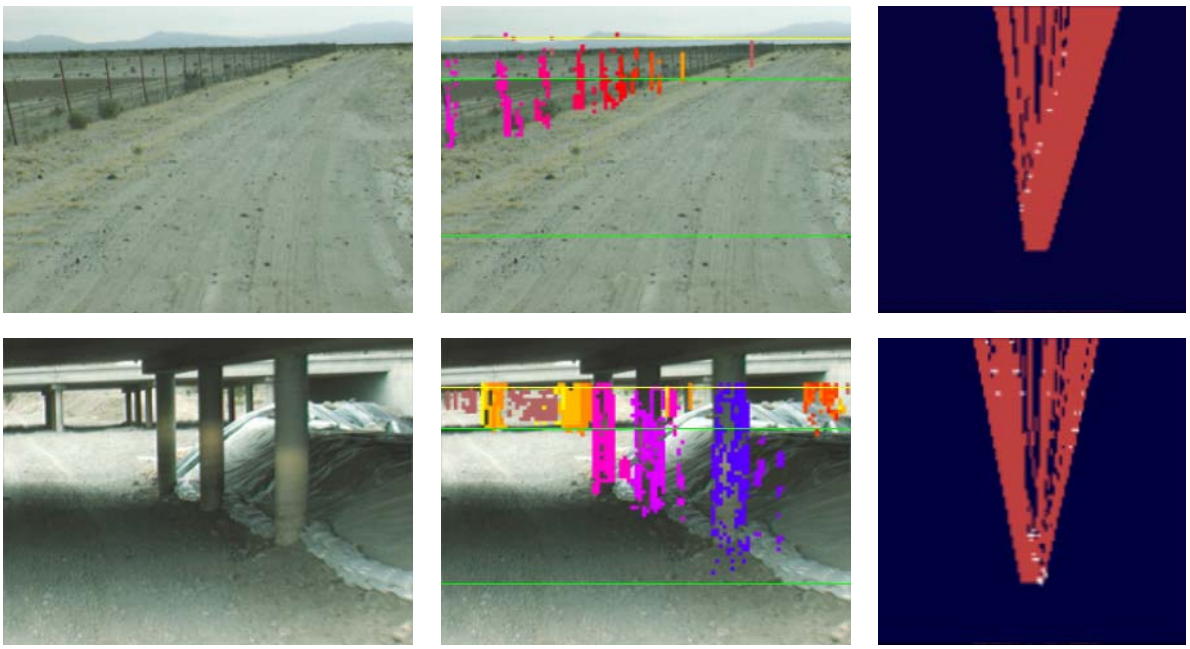
In case of a textureless obstacle, our algorithm can detect only its edges. This problem has not been addressed yet because of the computational weight of an image presegmentation [9] and because wide obstacles are seen easily by laser sensors.

5. Conclusions and Future Developments

The developed algorithm allows to extract in a robust way the cameras pitch angle at the time of acquisition from a pair of stereoscopic images acquired in unstructured environments. A straight line road profile assumption is made, but in the near future we aim to remove it and to extract the lateral road profile, as seen in [5]. In fact, as shown in fig. 12, in presence of a slope variation the ground correlation line is not a straight line, and contains the information about the road profile. Once the road profile is known, it will also be possible to better estimate the obstacles height in order to judge about their traversability.

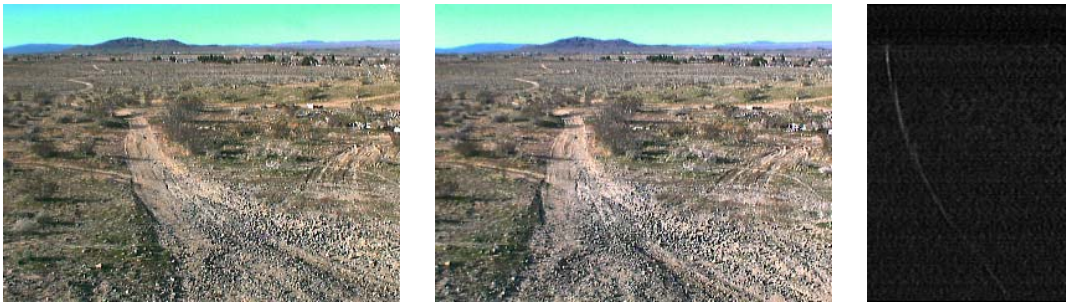
Furthermore, studying the ground correlation line discontinuities it may be possible to detect evident negative obstacles like cliffs in front of the vehicle (fig. 13).

Further improvements will be tested, in order to strengthen the algorithm against false positives and negatives, for example computing the DSI on each of the color channels separetely and comparing the results. By using the road profile information it will also be possible to obtain an accurate warped image [12] in order to perform a reliable road detection.



Original right images | Obstacle images | Bird's eye view maps

Figure 11. Mapping in world coordinates of fence posts and of an underpass. The right most images represent the maps of the 50m x 50m square region in front of the vehicle. Red indicates areas seen by the cameras, blue unknown areas (unseen or behind an obstacle), white detected obstacles.



Left image | Right image | V-disparity image

Figure 12. Slope changing effect in a V-disparity image

Left image | Right image | V-disparity image

Figure 13. V-disparity image in presence of a cliff

References

[1] P. Coulombeau and C. Laurgeau. Vehicle yaw, pitch, roll and 3D lane shape recovery by vision. In *IEEE Intelligent Vehicles Symposium, Versailles*, pages 646–651, June 2002.

[2] S. B. Goldberg, M. W. Maimone, and L. Matthies. Stereo vision and rover navigation software for planetary exploration. In *IEEE Aerospace Conference, Big Sky, Montana, USA*, volume 5, pages 2025–2036, Mar. 2002.

[3] K. Konolige. Small vision systems: Hardware and implementation. In *Eighth Intl. Symposium on Robotics Research, (Hayama, Japan)*, pages 111–116, Oct. 1997.

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

[5] R. Labayrade, D. Aubert, and J.-P. Tarel. Real time obstacle detection on non flat road geometry through V-disparity representation. In *IEEE Intelligent Vehicles Symposium, Versailles*, pages 646–651, June 2002.

[6] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry. In *Computer Vision and Pattern Recognition (CVPR) 2004*, volume 1, pages 652–659, June 2004.

[7] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *IEEE Workshop on Stereo and Multi-Baseline Vision, Kauai, HI*, Dec. 2001.

[8] S. Singh, R. Simmons, T. Smith, A. Stentz, V. Verma, A. Yahja, and K. Schwehr. Recent progress in local and global traversability for planetary rovers. In *IEEE International Conference on Robotics and Automation, San Francisco, USA*, Apr. 2000.

[9] J. Steele, C. Debrunner, and M. Whitehorn. Stereo images for object detection in surface mine safety applications. Technical Report TR20030109, Western Mining Resource Center, Colorado School of Mines, Nov. 2003.

[10] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, pages 323–334, Aug. 1987.

[11] W. van den Mark, F. Groen, and J.-C. van den Heuvel. Stereo based navigation in unstructured environments. In *IEEE Instrumentation and Measurement Technology Conference Budapest, Hungary*, 2001.

[12] T. A. Williamson. *A High-Performance Stereo Vision System for Obstacle Detection*. PhD thesis, Carnegie Mellon University, Sept. 1998.