Ground Plane Obstacle Detection with a Stereo Vision System

José A. Gaspar **José Santos-Victor** **João Sentieiro**

Instituto de Sistemas e Robótica,
Instituto Superior Técnico,
Av. Rovisco Pais, 1096 Lisboa Codex, PORTUGAL
e-mail gaspar@kappa.ist.utl.pt

Abstract. This paper presents a system for vision-based obstacle detection for mobile robots, developed within the framework of purposive vision.. A single camera and a set of flat mirrors are used to acquire two distinct images of the area in front of a mobile robot. The assumption that the robot is moving on the ground plane is explicitly used to constrain the perception problem. An obstacle is defined as any point laying outside of the ground plane.
During a calibration phase, a set of correspondent points are used to robustly estimate the set of 8 parameters that fully describe the disparity map of the ground plane. The disparity map, *Reference Map*, establishes the disparity for every image pixel, provided that the corresponding 3D point is on the ground plane. During operation, to detect the presence of obstacles, the system has to check for the points where the Reference Map is violated. A real time implementation was developed yielding good results.

1 Introduction

This paper addresses the problem of visual based mobile vehicle navigation, particularly at the obstacle avoidance level. Many of the developed systems rely on different sensors for obstacle avoidance, such as sonar, infrared, etc, but generally excluding vision, mainly due to the heavy computational load involved.

Traditionally, a computer vision approach would involve some sort of visual reconstruction, yielding a depth map which, in turn, would be used to determine the shape of obstacles and, possibly, compute a "safe" trajectory throughout the visible space. This approach has revealed itself extremely intensive, from the computational point of view and, beyond that, quite sensitive to uncertainty, and system parameters settings.

More recently, however, the approaches of *active vision* and *purposive vision* have contributed to significant improvements in these areas. The key idea is that the *observer* in a particular perception task should be *actively* involved in the task itself. Moreover, the visual algorithms should be designed in order to solve a particular problem, in contrast with the general reconstruction approach aiming at building a general map to solve any problem (navigation, manipulation, recognition, etc). This work fits into this framework. On one hand, we wish to use a high resolution sensing modality, such as vision, for obstacle avoidance; on the other hand, the algorithms will be specially tailored to this specific task in order to be fast and robust. No reconstruction is performed for this problem.

The system is based on a stereo algorithm for fast obstacle detection. Similarly to the work described in [6], an obstacle is defined as any object which does not lay on the ground plane (also known as GPOD - Ground Plane Obstacle Detection). With this constraint in mind, an offline procedure estimates a nominal disparity vector field, characterizing the ground plane. If there is an obstacle in the scene, then these nominal disparity vectors will be violated. Therefore, during normal operation, the system should just check for violations of the expected disparities, instead of fully reestimating these values, as a traditional system would do.

Relatively to the approach described in [6], this work differs in two main points. On one hand the stereo setup is based on a single camera and a mirror system, thus avoiding gain and aperture mismatches between a two camera setup, and leading to a smaller, more compact structure. On the other hand, the fact that the ground plane is flat is explicitly used, since the disparity vector field can be expressed as a function of 8 parameters. Accordingly, these parameters can be estimated from a reduced number of disparity measurements, and global consistency of the whole disparity vector field is guaranteed.

2 Setup

For the purpose being envisaged, the main approaches would either be based on optical flow (as in [8]), or on correspondence/disparity mechanisms (such as the stereo setup described in [6]). The optical flow approach uses a single camera mounted on a moving platform, but requires the knowledge of the camera trajectory. The robot motion used for the optical flow calibration (i.e. computing the flow corresponding to the ground plane) determines, to some extent, the permissible motion for current operation. The stereo approach, instead, can be tested independently of the robot trajectory (even in a static case), and does not exclude the extension to optical flow techniques. This points led to the choice of a stereo based technique.

The usual two camera setup for a stereo system has often the problem of mismatches between the camera gains, aperture, focal length, etc, which may harden the correspondence estimation problem. The setup used in this system is based on a single camera, hence overcoming the camera mismatch problems, and a system of 2 flat mirrors.

Even though this setup leads to a reduced field of view (unless convex mirrors are used), only one image acquisition needs to be performed each instant, with the right image half corresponding to one *virtual* camera and the left image half, to the other *virtual* camera. Furthermore, a compact robust mechanical structure can be obtained by suitable design. This setup is illustrated in Figure 1.

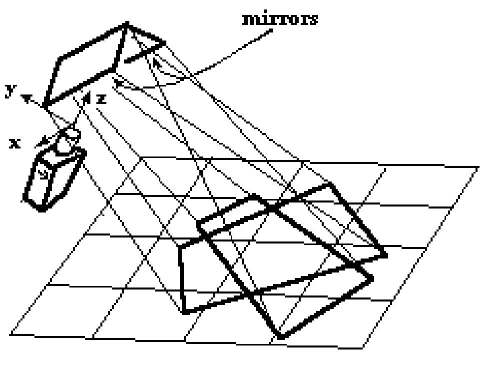


Fig. 1.: Experimental setup: a single camera and a set of flat mirrors are used to acquire stereo images of the space in front of a mobile robot.

3 GPOD Algorithm

The GPOD algorithm has two main steps: the computation of the reference map (initialization procedure), and the correspondences test during normal operation.

The reference map (RM) establishes for each pixel, (x_r, y_r), in the right image the expected ground matching pixel, (x_l, y_l), in the left image. The basic assumption is that the vehicle is moving on a ground plane, and any point outside this plane should be considered as an obstacle.

The ground plane constraint and the left-right camera coordinate transformation are given by:

$$[X_l\ Y_l\ Z_l\ 1]^T = T_r^l\ [X_r\ Y_r\ Z_r\ 1]^T \tag{1}$$

$$A.X_r + B.Y_r + C.Z_r + D = 0, \tag{2}$$

where (X_l, Y_l, Z_l) and (X_r, Y_r, Z_r) stand for the 3D coordinates of a given point, seen by the left or by the right camera, and A, B, C, D define the ground plane. Using a pin-hole camera model (perspective projection), the correspondences between left-right image points can be obtained in closed form:

$$\begin{bmatrix} x_l \\ y_l \end{bmatrix} = \mathcal{F}(x_r, y_r) \tag{3}$$

$$= \begin{bmatrix} f_x(x_r, y_r) \\ f_y(x_r, y_r) \end{bmatrix} \tag{4}$$

$$= \frac{1}{a_1 x_r + a_2 y_r + 1} \begin{bmatrix} b_1 x_r + b_2 y_r + b_3 \\ c_1 x_r + c_2 y_r + c_3 \end{bmatrix} \tag{5}$$

where $\mathcal{F}$ denotes the reference map (RM), and $u = [a_1\ a_2\ b_1\ b_2\ b_3\ c_1\ c_2\ c_3]^T$ is the parameter vector to be estimated in the initialization procedure.

Alternatively, the RM can be seen as a plane-to-plane projective transformation [10]:

$$\begin{bmatrix} \lambda x_l \\ \lambda y_l \\ \lambda \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \\ a_1 & a_2 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix} \tag{6}$$

To estimate u, we need at least 4 point correspondences of the ground plane. Let

$$P = \{(x_{l_1}, y_{l_1}, x_{r_1}, y_{r_1}), ..., (x_{l_N}, y_{l_N}, x_{r_N}, y_{r_N})\} \tag{7}$$

be a set of N reference point matches ($N \geq 4$). By writing equation (5) for the set of point matches in P and rearranging the terms, the following equation system is obtained:

$$Mu = v \tag{8}$$

where M and v depend solely on the elements of P. To improve robustness, however, an over constrained system is solved, and the least-squares solution for the parameter vector, u, can then be obtained through the pseudo-inverse solution:

$$u = [M^T M]^{-1} M^T v \tag{9}$$

or the simple inverse in the case of four data point matches.

An initial estimate of the RM parameters uses four patterns in the ground, which are matched manually. A corner detection algorithm is then applied to the input images, and the resultant corner points are matched automatically (the matching procedure being guided by the available parameterization), yielding the final reference correspondence map (see Figure 2).

During normal operation, the expected correspondences (explicit in the RM) are checked by calculating the "Sum of Square Differences" (SSD) [2] values for every pixel, (x, y), in the right image, (I_r), and the potential correspondent pixel in the left image, (I_l), (see Figure 3c). The SSD functional is given by

$$I_{SSDtest}(x,y) = SSD(\ I_l(\ \mathcal{F}(x,y)\ ),\ \ I_r(x,y)\ ) \tag{10}$$

$$= \sum_{i=-W}^{W} \sum_{j=-W}^{W} [\ I_l(f_x(x,y)+i, f_y(x,y)+j) - I_r(x+i, y+j)\ ]^2, \tag{11}$$

where W determines the SSD window size. A final step to detect ground plane constraint violations, consists in applying a threshold procedure to the computed SSD values. This procedure starts by applying a global threshold. Additionally, a second threshold is applied, based on the density of subsisting points in a given neighbourhood, thus eliminating noisy, sparse points.

4 Results

The algorithm presented is running on a 80486/50MHz PC computer. The operating frequency is about 0.6Hz, comprising image acquisition, obstacle detection and displaying, for images of 167 columns by 216 lines[1].

A disparity map computed during the calibration procedure is shown in Figure 2.

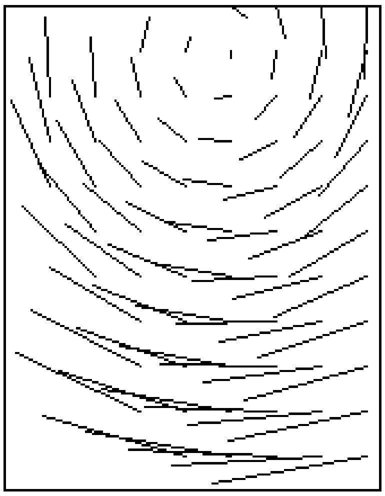
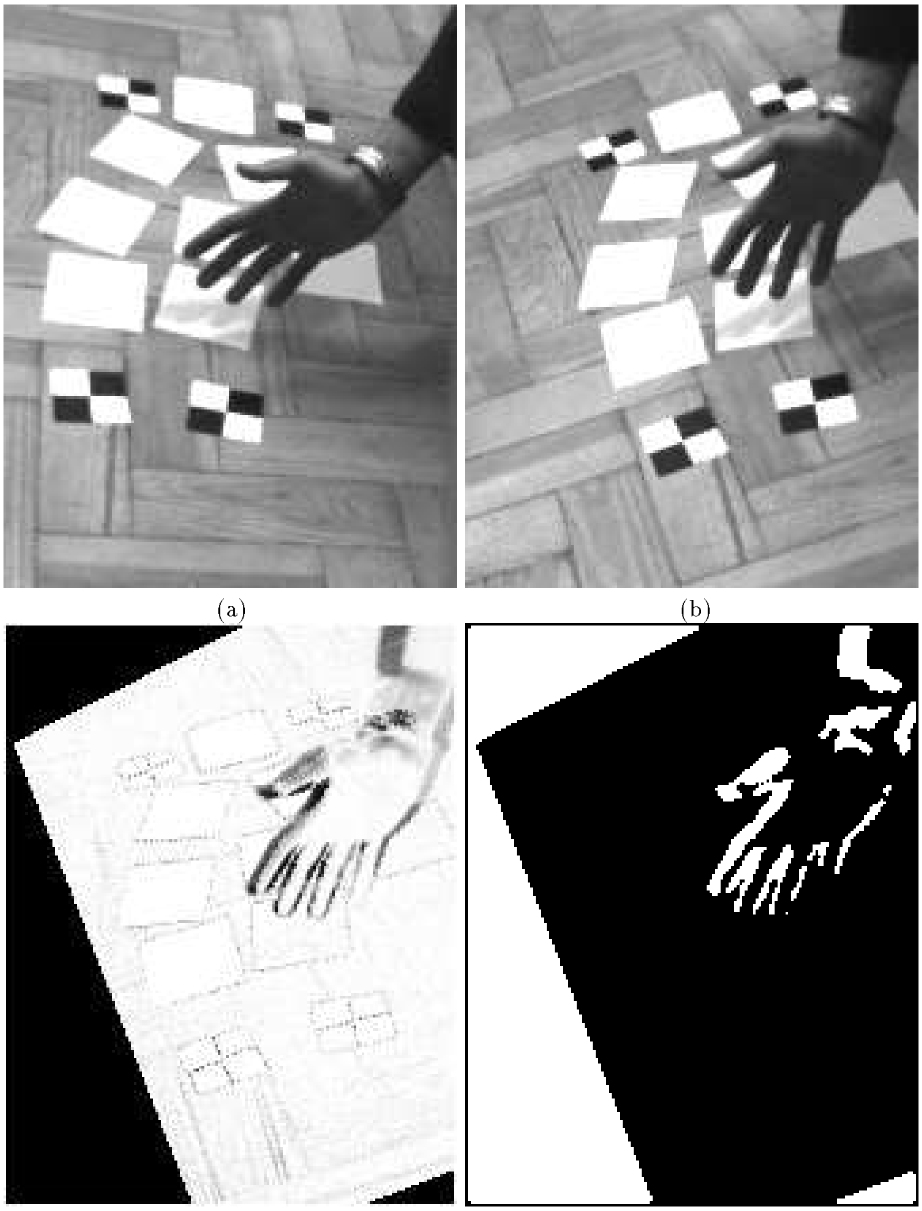


Fig. 2.: The reference disparity map contains the disparity of every right image pixel, provided that the corresponding 3D point lays on the ground floor.

The four patterns used for the initial RM parameterization can be seen in the images of Figures 3a and 3b.

[1] It strongly depends on the image and the SSD window sizes - for the results shown, the smallest window size (3×3) was used.



(a) (b) (c) (d)

Fig. 3.: Results: [a] left image (167 col. x 216 lines), [b] right image; [c] SSD test output (inverted) and [d]final output.

Figure 3 shows the acquired images (Figure 3a, Figure 3b), the SSD test output (Figure 3c) and the output image (Figure 3d), with the detected obstacles shown in white.

The textured areas of the obstacle (hand) are clearly detected, except in the case where the expected disparity is oriented along local edges, which is a known limitation in binocular stereo. However, these results are sufficient to implement simple reactive-type obstacle avoidance strategies for mobile robots.

5 Conclusions

This paper describes an algorithm for fast obstacle detection, suitable for mobile robots. The method relies upon the ground plane constraint, considering that every point outside the ground plane is an obstacle to be detected.

An experimental setup based on a single camera and a set of 2 mirrors was proposed in order to acquire a pair of stereo images. This setup not only avoids eventual parameter mismatches in a two-camera configuration but may also lead to a more robust and compact structure.

In the absence of obstacles, a set of point correspondences is determined, and used to estimate a set of 8 parameters which describe the correspondences between left and right images. Once in operation, the system will only check if the expected disparity is observed (for any point laying on the ground plane) or violated (for every other point outside the ground plane) thus detecting obstacles.

The parameterization used proved to be effective, easy to implement, and ensuring global consistency of the disparities over the whole visual field.

The system has been implemented and led to encouraging results, described in the paper. Future developments include further optimization to accelerate the overall speed and applying these techniques on a mobile robot to implement reactive-type obstacle avoidance mechanisms.

References

1. Y. Aloimonos, I. Weiss, A. Bandyopadhyay. Active Vision, *International Journal of Computer Vision*, 1(4):333-356, January 1988.
2. P. Anandan. A computational framework and an algorithm for the measurement of visual motion, *International Journal of Computer Vision*, 2(3):283-310, 1989.
3. R. Bajcsy. Active Perception vs. Passive Perception, Proceedings of IEEE workshop on Computer Vision, Bellair, MI, October 1985, pp. 55-59.
4. D. Ballard. Animate Vision, *Artificial Intelligence*, 48:57-86, 1991.
5. S. Cornell, J. Porril, J. Mayhew. GPOD under variable camera geometry using a predictive stereo matcher, BMVC 1992.
6. F. Ferrari, G. Garibotto, S. Masciangelo, G. Sandini. VOILA, ESPRIT Project P2502 (ECCV 1992 Esprit Open Day).
7. F. Ferrari, E. Grosso, G. Sandini, M. Magrassi. A stereo vision system for real time obstacle avoidance in unknown environment, IEEE international workshop on Intelligent Robots and Systems, IROS 1990.
8. F. Guarnotta, D. Rizzieri, F. Tarocchi, M. Tistarelli, G. Sandini. Studio de un problema di 'obstacle avoidance' per un robot mobile, DIST Genova, internal report.
9. H. Mallot, H. Bulthoff, J. Little, and S. Bohrer. Inverse perspective mapping simplifies opticla flow computation and obstacle detection. *Biological Cybernetics*, 64:177–185, 1991.
10. J.L. Mundy and A. Zisserman (editors). Geometrical Invariance in Computer Vision. MIT Press, 1992.

This article was processed using the TEX macro package with IRS94 style