Stereo Vision System Applied to the Reconstitution of Polyhedric Structures

Cristiano Jacques Miosso, Mirele de Almeida Mencari and Adolfo Bauchspiess

Electrical Engineering Department–ENE
Technology Faculty –FT
University of Brasília – UnB, BRAZIL
Miosso@engineer.com, bauchspiess@ene.unb.br

Abstract — **Stereo machine vision's purpose is to determine spatial coordinates of opaque surfaces in a scene by joint processing of images taken by cameras in different positions. In this paper a binocular vision system is implemented based on the matching of strategic points in images, which leads to sparse depth maps. Only the detected edges are rectified. The implemented algorithms have been used for the reconstitution of five different polyhedrons. The calibration procedure was found to be the most critical step with regard to the achievable precision.**

Keywords: **stereo machine vision, camera calibration, stereo matching, photogrammetry.**

I. INTRODUCTION

An important application of digital image processing in the area of control and automation is robot guidance. It is to determine, starting from the captured images of an environment, the coordinates of a trajectory to be followed, of objects to be manipulated or of obstacles to be avoided. For this purpose, the use of a single image of the scene is not sufficient because the depth information is lost.

A solution commonly used is the employment of active sensors, such as infrared, laser or ultra-sound, that allow to recover the depth information. But, in general, they are very restrictive, because only the depth of a single point is obtained.

With stereo vision, the technique used in this paper, the whole scene can be analyzed to obtain the 3D spatial coordinates necessary to perform the desired task.

II. IMAGE FORMATION PROCESS AND 3D VISION

A. Perspective Projection

In this paper the *pinhole* camera model will be used, which assumes well focused images, i.e., objects are restricted to the depth of field of the employed optical system. The main elements are the retina plane *P* and the optical center *C*. The image *M´* of a point *M* is the intersection of *P* with the light ray that contains M and C (fig. 1).

The image formation consists thus of a perspective projection. Points localized in the 3D space are projected in the retina plane by

$$\mathbf{r'} = \frac{z'}{\mathbf{r} \cdot \hat{\mathbf{z}}} \mathbf{r} ,\qquad(1)$$

where $\hat{\mathbf{z}}$ is the unitary vector at the *z* axis and **r** and **r'** are the positions of the object and its image with respect to C. Points in the focal plane *F* are projected on *P* at infinity.
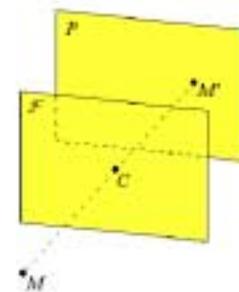

Fig. 1. Perspective projection in a *pinhole* camera.

B. 3D Vision Techniques

Different 3D vision techniques are available for the analysis of three-dimensional objects from their two-dimensional images. Some just determine the 3D shape of objects, without obtaining measurements of the object. Photogrammetry techniques, on the other hand, provide cartesian coordinates *x,y,z* of the surfaces, so that shapes, positions and dimensions can be reconstituted.

To the first group belongs the photometric stereo vision [5]. Different images of a scene are formed, in this case, for a camera in a fixed position, each one of them under specific and controlled conditions of illumination. From the brightness differences in the images in each case, it is then possible to determine the orientation of normal lines to the analyzed surface, and thus to evaluate its form.

Of more complex implementation is the technique *shape from shading* [5]. It is to determine, starting from a single image, the form of the constituent surfaces, using the irradiation differences at the retina plane. Although it involves more complex algorithms, there is the advantage of using a single image, and it is not necessary to control the illumination.

Both mentioned techniques make use of the luminance information or color in the images, and not just the simple geometric relationships associated to the perspective projection. A more restricted approach to exploit these relationships is binocular stereo vision, the technique employed in this paper. Its basic idea is to use two or more images of the scene to be analyzed, taken by cameras in different positions. Unlike the previous techniques, the binocular vision approaches the photogrammetric problem, and has the advantage of not demanding a rigorous control of the illumination as for the photometric vision, and of being, in its foundation, simpler than the *shape from shading* technique.

Fig. 2 illustrates how the use of more than one image allows to resolve the ambiguity inherent to one single

image. The position of one object P can be unequivocally determined from its images P_1 and P_2. This figure illustrates the fact that trough the intersection of the optical rays associated with each point that forms a surface it is possible to reconstruct it completely, that is, not only the shape but also the positions and dimensions of the represented objects. The stereo vision technique allows, so to speak, to invert the perspective projection process that happens when the images are formed.
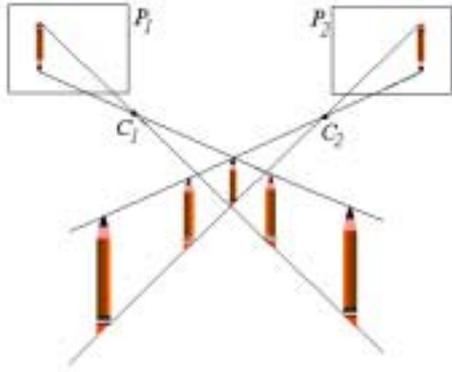

Fig. 2. The use of binocular stereo vision to determine the position of an object P starting from its images P'_1 and P'_2

The use of binocular stereo vision leads, however, to two difficulties of practical order. In first place, the evaluation of all the necessary characteristics of the camera, constituted of its intrinsic and extrinsic parameters (position and orientation in relation to the external coordinate system) is denominated *calibration*, and constitutes one of the central problems in stereo vision. The second critical aspect is with respect to the determination of the homologue points in the different images, that is, of those that correspond to projections of a same point in the surface of the represented object.

III. Methodology

A. *Generic description of the Implemented System*

As there is not a general solution for the correspondence problem, we restricted the application to the reconstitution of simple polyhedric objects, well represented by straight line segments (edges).

Fig. 3 presents the block diagram that describes the implemented binocular vision system. Initially, two images of the solid are taken, from different positions. The stereo images are then submitted to a border detection algorithm, originating two binary images that represent the edges of the polyhedron. These binary images are then rectified, with the objective of aligning the homologoue points in the two images and thus to simplify the following stage, the correspondence. The obtained correspondence map allows finally to reconstitute the space coordinates of the object.

B. *Image Acquisition*

For the evaluation of the implemented algorithms, we built four prisms of different bases, as well as a pattern of thirty two points on two orthogonal planes, for the calibration of the optical system. Images of these objects were acquired by a single monochrome CCD camera,

alternated between two positions. All images were transmitted to a PC by a OCULUS frame-grabber controlled by the driver ODTCX, and stored in TIFF format by the ODCI application. The algorithms were implemented off-line with MatLab 5.0.
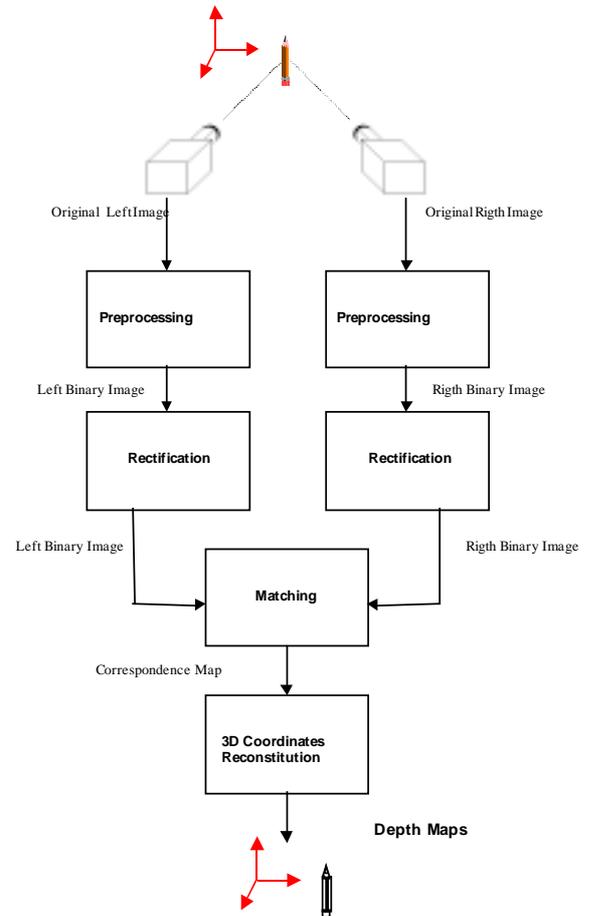

Fig. 3. Block diagramm of the stereo vision system

C. *The Calibration of the Optical System*

By means of the photogrammetry we aim to determine the space coordinates of surfaces with regard to an external, pre-determined system. For this purpose, it is necessary to obtain a relationship similar to (1), but where the coordinates of P are referenced to that system. The position of P', for its time, should be specified by the indexes (line and column) of the corresponding pixel in the retina plane, because digital images are always specified with regard to these values, and not as distances to the principal point or the optical center.

This relationship is determined by the calibration matrix of the optical system, designated by $\tilde{\mathbf{P}}$. Where x, y, z are the coordinates of a point P in relation to an arbitrary system, and i and j are the indexes of the corresponding *pixel* in the image P', we have [2]:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \tilde{\mathbf{P}} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \qquad (2)$$

where $j = \dfrac{U}{S}$ and $i = \dfrac{V}{S}$.

The calibration matrix, or of perspective transformation, is function of parameters that depend exclusively on the camera and of the external system of coordinates. The first set, denominated *intrinsic parameters*, comprehends the distance of the optical center to the retina plane, f, the coordinates of the principal point (i_0, j_0), and the distances δu and δv of a pixel to its neighbors in the horizontal and in the vertical, respectively. The other set, denominated *extrinsic parameters*, include the position C of the optical center in relation to the external system and the unitary vectors $\hat{\mathbf{h}}$, $\hat{\mathbf{v}}$ and $\hat{\mathbf{a}}$ that describe the orientation of the camera in relation to this system, according to Fig. 4.

In function of these variables, we have:

$$\tilde{\mathbf{P}} = \begin{pmatrix} \dfrac{f}{\delta u} & 0 & j_0 \\ 0 & \dfrac{f}{\delta v} & i_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \left(\begin{array}{c|c} \hat{\mathbf{h}}^T & -\mathbf{C}\cdot\hat{\mathbf{h}} \\ \hat{\mathbf{v}}^T & -\mathbf{C}\cdot\hat{\mathbf{v}} \\ \hat{\mathbf{a}}^T & -\mathbf{C}\cdot\hat{\mathbf{a}} \end{array} \right), \qquad (3)$$

where the two matrixes depend exclusively on intrinsic and extrinsic parameters, respectively.
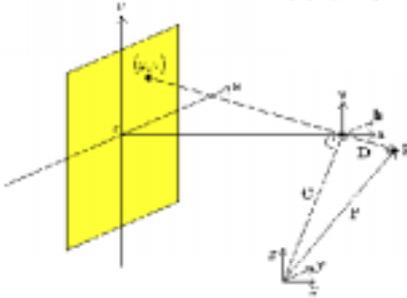

Fig. 4. Vectors that define position and oritentation of the camera with respect to the external coordinates system

The reconstitution of a scene from two stereo images demands that the perspective transformation matrixes $\tilde{\mathbf{P}}_1$ and $\tilde{\mathbf{P}}_2$ should be determined.

Some variables that determine the calibration matrix cannot, however, be directly measured [4]. Although δu and δv are usually supplied by the camera manufacturer, on the distance f is just known to vary between two well-known values, but not its exact value. The position of the optical center inside the camera is also unknown, what makes unfeasible the measurement of vector C. The difficulty in determining the orientation of the optical axis of the camera accurately with regard to the external system impedes the acquisition of $\hat{\mathbf{h}}$, $\hat{\mathbf{v}}$, $\hat{\mathbf{a}}$.

The calibration in the two positions A and B shown in Fig. 3 was done using the image of N reference points whose positions in relation to the adopted system of coordinates are known; next, trough the resolution of a linear system we calculate the matrix that maps those points in the positions (i, j) of the corresponding pixels in the image.

Being (x_m, y_m, z_m) the coordinates of the m-th point and (i_m, j_m) the indexes of the corresponding *pixel* in the image obtained by the calibration matrix $\tilde{\mathbf{P}}$, we have from (2):

$$\begin{cases} j_m = \dfrac{\tilde{\mathbf{P}}_{11} \cdot x_m + \tilde{\mathbf{P}}_{12} \cdot y_m + \tilde{\mathbf{P}}_{13} \cdot z_m + \tilde{\mathbf{P}}_{14}}{\tilde{\mathbf{P}}_{31} \cdot x_m + \tilde{\mathbf{P}}_{32} \cdot y_m + \tilde{\mathbf{P}}_{33} \cdot z_m + \tilde{\mathbf{P}}_{34}} \\[4mm] i_m = \dfrac{\tilde{\mathbf{P}}_{21} \cdot x_m + \tilde{\mathbf{P}}_{22} \cdot y_m + \tilde{\mathbf{P}}_{23} \cdot z_m + \tilde{\mathbf{P}}_{24}}{\tilde{\mathbf{P}}_{31} \cdot x_m + \tilde{\mathbf{P}}_{32} \cdot y_m + \tilde{\mathbf{P}}_{33} \cdot z_m + \tilde{\mathbf{P}}_{34}} \end{cases}. \qquad (4)$$

A system of *2N* equations is obtained, whose unknowns are the 12 elements of $\tilde{\mathbf{P}}$. Although 6 points are sufficient, 32 white squares distributed on 2 perpendicular dark planes were used to give more robustness to the calculation.

D. Stereo Images Correspondence

There are three kinds of techniques commonly used in determining stereo images correspondence, in machine vision systems [5, chapter 13]. The first of them, called *feature matching*, is based on the extraction from each image, separately, of some feature which generates a simple symbolic description and which thus allows a more direct association with the other images. This feature may be, for example, the set of points for which the luminance in the images has nonzero Gaussian curvature. The most general and most widely used of them is, however, the set of points in the edges of each image [7].
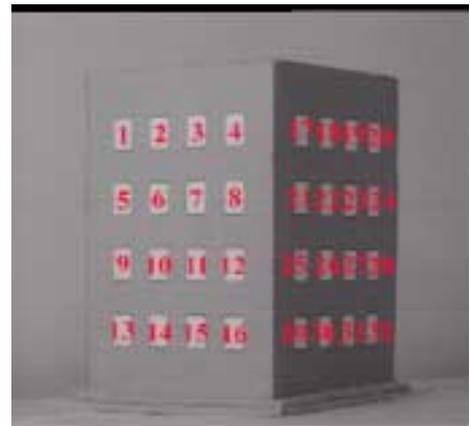

Fig. 5. Pattern with 32 reference points in 2 orthogonal planes, used during calibration of a CCD camera in positions A and B.

In this research, we chose to apply this technique, based on edge detection in the first stage of the process. This is the approach which has led to the best results so far [5], and, although it gives rise to sparse depth maps, it allows a considerable reduction in computational effort since only points extracted in the first stage are analyzed during definition of the correspondence pairs.

1 Edge Detection in the Stereo Images

Edge detection's aim is to determine the imaged object's frontiers through automatic processing of the color or gray level (luminance) information available in each pixel. This procedure has many applications in image processing and computer vision, and is an indispensable technique in both biological and machine vision systems [6].

Algorithms based on specific linear and shift-invariant filters constitute the most widely used procedure to

extract edges in digital images, and the one with less computational effort associated. In the case of first-order filters, an edge is interpreted as an abrupt variation in gray level between two neighbor *pixels*.

2 Rectification of Stereo Images

In this research, the algorithm used during the rectification process was that presented in [1]. It's based on the computation of two projection matrices $\tilde{\mathbf{P}}_{n1}$ and $\tilde{\mathbf{P}}_{n2}$ which describe the canonic system with the same optical centers than those of the original system used to produce the stereo images, described by known projection matrices $\tilde{\mathbf{P}}_1$ and $\tilde{\mathbf{P}}_2$. $\tilde{\mathbf{P}}_{n1}$ and $\tilde{\mathbf{P}}_{n2}$ alone are enough to compute the rectified images.

In fact, if $\tilde{\mathbf{P}}_o = \left(\mathbf{P}_o \middle| \mathbf{p}_o\right)$ and $\tilde{\mathbf{P}}_n = \left(\mathbf{P}_n \middle| \mathbf{p}_n\right)$ are the perspective transformation matrices before and after rectification, respectively, we have:

$$\begin{bmatrix} U_n \\ V_n \\ S_n \end{bmatrix} = \lambda \cdot \mathbf{P}_n \cdot \mathbf{P}_o^{-1} \cdot \begin{bmatrix} U_o \\ V_o \\ S_o \end{bmatrix}, \tag{5}$$

where λ is a constant and $(i_n, j_n) = \left(\dfrac{V_n}{S_n}, \dfrac{U_n}{S_n}\right)$ and $(i_o, j_o) = \left(\dfrac{V_o}{S_o}, \dfrac{U_o}{S_o}\right)$ are the pairs of coordinates of pixels in rectified and original images, respectively.

Equation (**5**) allows to compute the position of the image of a punctual object in each rectified retina plane given its image in the corresponding plane in the original configuration.

The rectification problem is then reduced to the computation of the rectifying matrices $\tilde{\mathbf{P}}_{n1}$ and $\tilde{\mathbf{P}}_{n2}$.

3 Determining the Homologue Pairs in the Stereo Images

The technique here adopted to solve the matching problem consists in a pre-processing stage, in which edge detection in the stereo images takes place, followed by rectification of the binary images obtained. The homologue pairs are then determined through sweeping of all horizontal lines in the rectified images. Since the homologue points are always in the same line, after rectification, an association of points in common lines in both images can be made (as Fig. **3** shows, the position of the polyhedrons with respect to positions A e B guarantees that their edges appear in the same sequence in a horizontal line in both images). Fig. **7** depicts this process.

However, a problem arises when the number of points in a horizontal line in one of the images does not match that found in the other image. In this case, we chose to eliminate the corresponding line in this situation.

Notice that this algorithm has as inputs two binary stereo images and as output a sparse matching map. Since the system was restricted to the analysis of polyhedric structures, however, this map allowed the reconstitution not only of points in the detected edges but also of internal ones. All was necessary to do was to compute

lines that connect points in those edges which are in the same horizontal line, in each rectified image. Trough this procedure, epipolar lines were reprojected in the tridimensional space, which allowed a reconstitution of points which are not present in the matching maps.
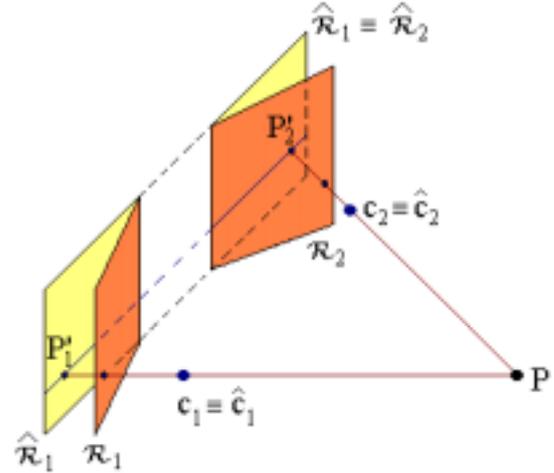
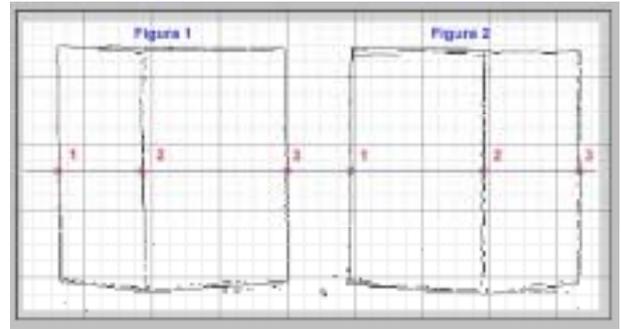Fig. 6. Transformation of the retina planes of two stereo images, during the rectification process

Fig. 7. Automatic matching of edges in the rectified stereo images

E. Reconstitution of the Polyhedrons

After determining pairs of homologue points in the edges of each polyhedron's images, the positions of the corresponding points in the scene were computed using the intersection of the light rays from these to the optical centers, as it is depicted in Fig. 2-a. This result may also be represented algebraically, according to the following analysis.

If (i_1, j_1) and (i_2, j_2) are homologue points in two stereo images formed by the optical systems $\tilde{\mathbf{P}}_1$ and $\tilde{\mathbf{P}}_2$, we obtain from (2) the following equations, in homogeneous coordinates:

$$\begin{bmatrix} U_k \\ V_k \\ S_k \end{bmatrix} = \tilde{\mathbf{P}}_k \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \tag{6}$$

where $j_1 = \dfrac{U_1}{S_1}$, $i_1 = \dfrac{V_1}{S_1}$, $j_2 = \dfrac{U_2}{S_2}$, $i_2 = \dfrac{V_2}{S_2}$ and $[x\ y\ z]^T$ is the position (to be computed) of an object which images are (i_1, j_1) and (i_2, j_2).

579

Making $\mathbf{w} = [x\ y\ z]^T$, we have:

$$j_1 = \frac{\mathbf{a}_1^T \cdot \mathbf{w} + a_{14}}{\mathbf{a}_3^T \cdot \mathbf{w} + a_{34}}\ ,\ i_1 = \frac{\mathbf{a}_2^T \cdot \mathbf{w} + a_{24}}{\mathbf{a}_3^T \cdot \mathbf{w} + a_{34}}\ ,\qquad (7)$$

$$j_2 = \frac{\mathbf{b}_1^T \cdot \mathbf{w} + b_{14}}{\mathbf{b}_3^T \cdot \mathbf{w} + b_{34}}\ ,\ i_2 = \frac{\mathbf{b}_2^T \cdot \mathbf{w} + b_{24}}{\mathbf{b}_3^T \cdot \mathbf{w} + b_{34}}\ ,\qquad (8)$$

where:

$$\widetilde{\mathbf{P}}_1 = \begin{pmatrix} \mathbf{a}_1^T & a_{14} \\ \mathbf{a}_2^T & a_{24} \\ \mathbf{a}_3^T & a_{34} \end{pmatrix} \text{ and } \widetilde{\mathbf{P}}_2 = \begin{pmatrix} \mathbf{b}_1^T & b_{14} \\ \mathbf{b}_2^T & b_{24} \\ \mathbf{b}_3^T & b_{34} \end{pmatrix}.\qquad (9)$$

Equations (7) and (8) can be rewritten, in matrix form:

$$\begin{pmatrix} \mathbf{a}_1^T - j_1 \cdot \mathbf{a}_3^T \\ \mathbf{a}_2^T - i_1 \cdot \mathbf{a}_3^T \\ \mathbf{b}_1^T - j_2 \cdot \mathbf{b}_3^T \\ \mathbf{b}_2^T - i_2 \cdot \mathbf{b}_3^T \end{pmatrix} \cdot \mathbf{w} = \begin{pmatrix} j_1 \cdot a_{34} - a_{14} \\ i_1 \cdot a_{34} - a_{24} \\ j_2 \cdot b_{34} - b_{14} \\ i_2 \cdot b_{34} - b_{24} \end{pmatrix}.\qquad (10)$$

The solution of this system, usually over-determined, leads to the coordinates $\mathbf{w} = [x\ y\ z]^T$ needed.

IV Results

The implemented stereo machine vision was applied to the reconstitution of the 3D-coordinates of a cube and of two prisms (with pentagonal and hexagonal bases, respectively). Fig. 8 and 9 depict the acquired stereo images, the detected edges and a bidimensional representation of x, y, z axes with all reconstituted points, for two cases (cube and prism with pentagonal basis). The results are compared to the known models of the analyzed structures.

Since the most critical stage of the whole algorithm is the automatic stereo matching, the reconstitution of the polyhedrons' vertices was also carried on based on manual matching of the stereo images. It was then possible to evaluate the results of the optical systems' calibration and to determine when experimental errors were due to wrong matching or to imprecision during the calibration procedure or image acquisition.

We observe that the known models of the polyhedrons were not completely covered by the reconstituted points. This is due to the fact that matching of stereo images was only accomplished when the horizontal epipolar lines intersected the two images' edges in the same amount of points. All points eliminated during this process are associated to the regions which are not covered by the reconstituted images.

The adoption of sparse depth maps shouldn't interfere, however, with estimation of distances from the camera to each analyzed polyhedron. So, table 1 shows the distances from each of them to the optical centers of the camera in both positions adopted.

TABLE 1
DISTANCES FROM THE POLYHEDRONS TO THE CAMERA'S OPTICAL CENTERS IN POSITIONS A AND B, ACCORDING TO THE IMPLEMENTED STEREO VISION SYSTEM (REAL VALUE OF 713mm FOR POSITION A AND 714mm FOR POSITION B)

Object	Distance to optical center in A (mm)	Distance to optical center in B (mm)
Calibration pattern	712,99	713,78
Cube	710,68	710,60
Prism with pentagonal basis	700,35	697,83
Prism with hexagonal basis	702,69	699,33

TABLE 2
CAMERA'S EXTRINSIC PARAMETERS OBTAINED DURING CALIBRATION PROCEDURE, FOR POSITIONS A AND B.

	Measured valor for camera in position A			Measured valor for camera in position B		
	x (mm)	Y (mm)	z (mm)	x (mm)	y (mm)	z (mm)
A axis	0,588	0,803	-0,096	0,810	0,580	-0,091
V axis	-0,032	-0,095	-0,995	-0,070	-0,059	-0,996
H axis	0,808	-0,588	0,0296	0,583	-0,813	0,0066
Optical Center (C)	-380,03	-600,84	54,03	-594,97	-391,52	46,98

V. CONCLUSION

The results obtained with the implemented algorithms were satisfactory. The reconstituted points of three test prisms matched well with the known model. The distance of the camera to the objects were estimated with errors under 3%, which is quite sufficient for robot navigation.

A restriction of the proposed approach is that it only works well with objects that can be adequately characterized by their borders. But very often this is the situation when a robot navigates in a structured environment.

V. REFERENCES

[1] Fusiello, A., E. Trucco and A. Verri (1998). Rectification with unconstrained stereo geometry.

[2] Grewe, L. L. and A. C. Kak (1994). Stereo Vision. Robot Vision Laboratory, School of Electrical Engineering, Purdue University, West Lafayette, Indiana.

[3] Koschan, A. and V. Rodehorst (1995). Towards Real-Time Stereo Employing Parallel Algorithms For Edge-Based And Dense Stereo Matching. Proc. of the IEEE Workshop on Computer Architectures for Machine Perception CAMP'95, 18–20, September, Como, Italy.

[4] Krick, F. (1995). *The Astonishing Hypothesis*. Touchstone, Simon & Schuster Inc., New York.

[5] Horn, B. K. P. (1986). *Robot Vision*. MIT Electrical Engineering and Computer Science Series. The MIT Press, McGraw-Hill Book Company, Cambridge, Massachusetts.

[6] Iwahori, Y., Md. S. Bhuiyan and A. Iwata (19--). Optimal edge detection under difficult imaging conditions. Technical report, Educational Center for Information Processing and Dept. of Electrical and Computer Engineering. Nagoya Institute of Technology, Showa, Nagoya, 466-8555, URL: http://www.center.nitech.ac.jp /people/bhuiyan/pub.html.

[7] Redert, A., E. Hendriks and J. Biemond (1999). Correspondence estimation in image pairs — designing estimators for dense geometric correspondence fields. *IEEE Signal Processing Magazine — 3D and Stereoscopic*, Visual Communication (Maio), Vol. 16, nº 3, pp 29–46.

[8] Thompson, C. M. and L. Shure (1995). *Image Processing Toolbox User's Guide — For Use With MatLab*. The MathWorks, Inc., January 1995.

[9] Torreão, J. R. A. (1999). 3D shape estimation in computer vision. *Controle & Automação*, Vol. 10, nº 2, pp 118–123.

APPENDIX

Results of the Reconstitution of Polyhedrons

(a) Stereo Image, (b) Reconstituted Points from manual correspondence in light gray, and known model of the surface in dark gray, (c) Detected borders, (d) Reconstituted Points in automatic correspondence modus in dark gray and model in light gray.
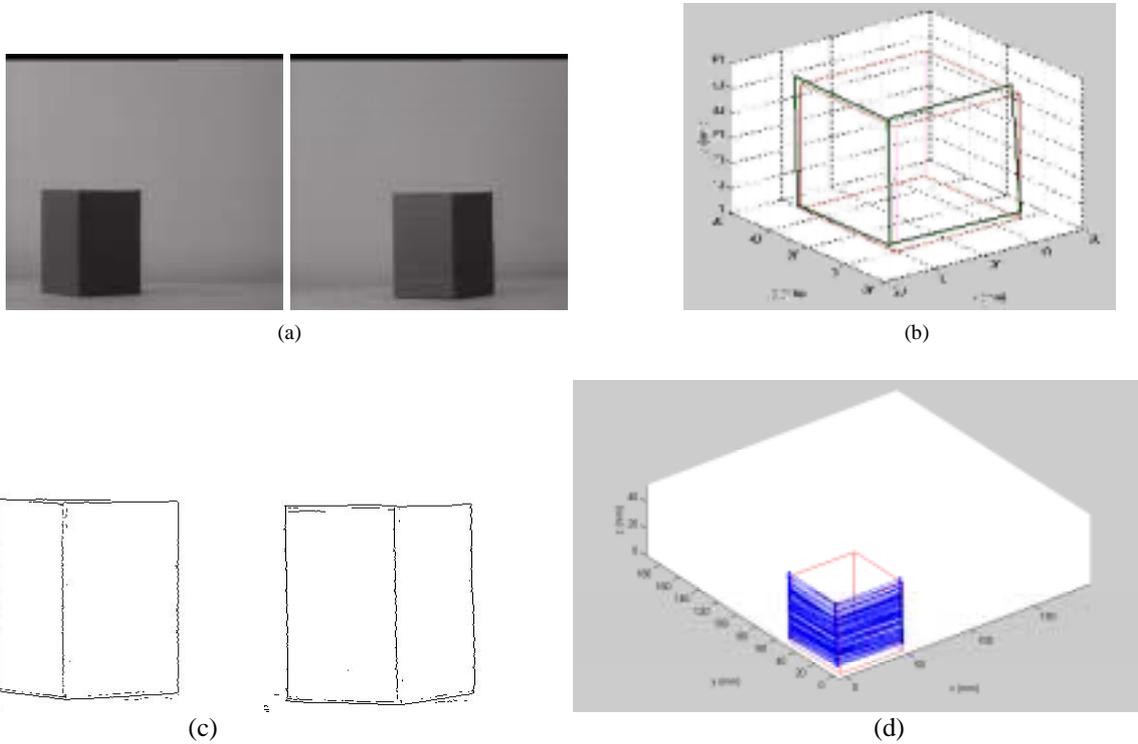

(a)

(b)

(c)

(d)

Fig. 8: Reconstitution of space coordinates of a cube, in manual and automatic correspondence of a stereo image pair.
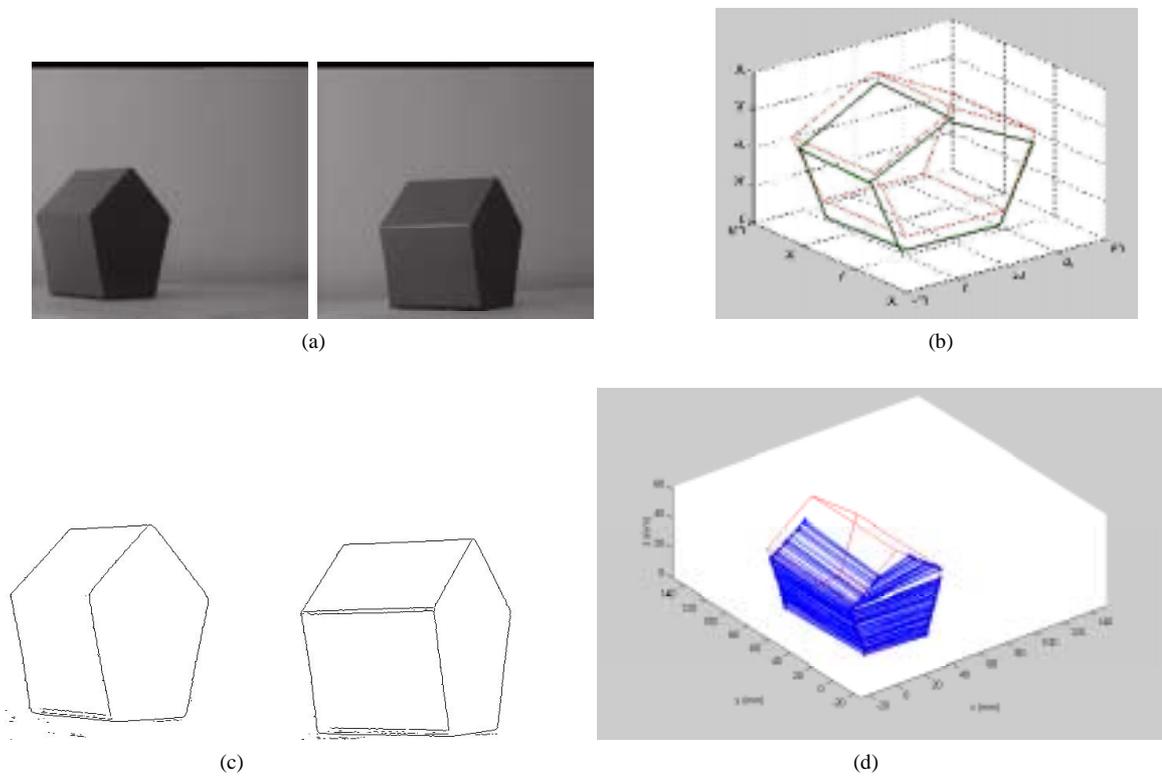

(a)

(b)

(c)

(d)

Fig. 9: Reconstitution of space coordinates of a pentagonal base prism, in manual and automatic correspondence of a stereo image pair.