3D PHOTOGRAMMETRY OF POLYHEDRAL STRUCTURES USING FUZZY BORDER DETECTION

Cristiano Jacques Miosso*   Adolfo Bauchspiess**

Departamento de Engenharia Elétrica — ENE
Faculdade de Tecnologia — FT
Universidade de Brasília — UnB
*miosso@engineer.com
*adolfo@ene.unb.br

Abstract: The purpose of photogrammety using stereo vision is to determine spatial coordinates of opaque surfaces by joint processing images taken from different positions. A binocular system is implemented which is based on matching strategic points in stereo images, leading to sparse depth maps. In a preprocessing stage a fuzzy edge detector is employed, reducing in approximately 90% the number of pixels used in the following computations. Only the detected edges are rectified. The algorithms we re applied to the reconstitution of poly hedrons. Calibration procedure was found to be the most critical step regarding the achievable precision. Copyright © 2001 IFAC.

Keywords: Stereo vision, image matching, fuzzy systems, robot navigation.

1. INTRODUCTION

Robot guidance is an important application of image processing in control and automation. The problem consists in determining coordinates of trajectories to be followed, of objects to be manipulated or of obstacles to be avoided, based on images taken from the environment. A single image is not enough if only geometric relations between objects and their images positions are to be considered, for depth information is lost in the perspective projection that happens when images are formed.

An often used solution is based on active sensors, such as infrared, laser or ultra -sound, which allow depth information to be recovered. But in general it is very restrictive, for a single point's depth is obtained.

The stereo vision technique used in this paper al lows the whole scene to be analyzed so all 3D spatial coordinates necessary to the task are computed.

2. IMAGE FORMATION PROCESS AND 3D VISION

2.1 Perspective Projection

In this paper the *pinhole* camera model will be used, which assumes well focused imag es, i.e., objects are restricted to the optical system's depth of field. The main elements are the retina plane *P* and the optical center *C*. A point M's image, *M´*, is the intersection of *P* with the light ray containing *M* and *C* (figure 1).

Image formation pr ocess consists thus of a perspective projection. Points in 3D space are projected in the retina plane according to

$$\mathbf{r}' = \frac{z'}{\mathbf{r} \cdot \hat{\mathbf{z}}} \mathbf{r}\,, \tag{1}$$

where $\hat{\mathbf{z}}$ is the unitary vector at *z* axis and $\mathbf{r}$ and $\mathbf{r}$' are object's and image's positions with respect to *C*.

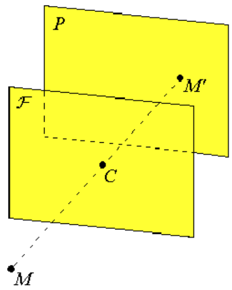

Fig. 1. Perspective projection in a *pinhole* camera.

2.2 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 them. On the other hand, photogrammetry techniques provide surfaces' cartesian coordinates, so shapes, positions and dimensions can be reconstituted.

The first group includes *photometric stereo vision* (Horn, 1986). Different images from the scene, each under specific and controlled illumination conditions, are furnished by a camera in a fixed position. From the brig htness differences in the images, it is then possible to determine the orientation of normal lin es to the analyzed surface, and thus to evaluate its form.

The *shape from shading* technique (Horn,1986) is more complex. Its aim is to determine the form of the surfaces based on a single image, using the irradiation differences at the retina plane. Despite the greater algorithms' complexity, they employ a single image, and there is no need to control illumination.

Both mentioned techniques use images' luminance information, rather than relying just on the simple geometric relationships associated to perspective projection. The stereo vision technique employed in this paper is restricted to these relationships. Its basic idea is to use two or more images from the scene to be analyzed, taken by cameras in different positions. Unlike the previous approaches, binocular vision attacks the photogrammetric problem; it does not depend on a rigorous control of illumination as photometric vision does, and it is, in its basis, simpler than the shape from shading technique.

Figure 2 shows how using more than one image allows the ambiguity inherent to a single image to be solved. The position of an object *P* can be unequivocally determined from its images P_1 and P_2. This figure illustrates the fact that through the intersection of the optical rays associated to each point in a surface it is possible to reconstruct it completely, not only the shape but also the positions and dimensions. The stereo vision technique allows, so to speak, to revert the perspective projection process that happens when the images are formed.

Binocular stereo vision leads, however, to two implementation problems. In first place, the evaluation of all the necessary characteristics of the camera, including its intrinsic and extrinsic parameters (position and orientation in relation to the external coordinate system) constitutes one of the central problems in stereo vision, called *calibration*. The second critical aspect relates to determining the homologue points in the different images (those that correspond to projections of a same point in the surface of the represented object).

3. METHODOLOGY

3.1 Generic description of the Implemented System

Since there is no general solution to the matching problem, application was restricted to polyhedrons, well represented by straight line segments (edges).

Figure 3 presents a block diagram describing 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 polyhedron's edges. These binary images are then rectified, so that homologue points in the two images are aligned. This simplifies the following stage — the matching process. The resulting correspondence map then allows the space coordinates of the object to be to reconstituted.

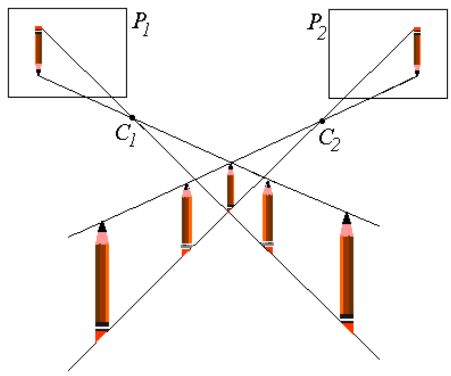

Fig. 2. Binocular stereo vision used to determine the position of an object *P* from its images P'_1 and P'_2.

3.2 Image Acquisition

To evaluate the algorithms, 4 prisms with different bases were built, as well as a pattern with 32 points on two orthogonal planes, used in the optical system calibration. Images from these objects were taken by a single monochrome CCD camera, altered between two positions. All images were transmitted to a PC by an OCULUS frame-grabber controlled by driver ODTCX, and stored in TIFF format by ODCI. The algorithms were implemented off-line in MatLab.

3.3 Optical System Calibration

Photogrammetry's aim is to determine surfaces' space coordinates 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'*, in turn, should be specified by the indexes (line and column) to the corresponding pixel in the retina plane, since information in digital images is always addressed by these values.

This relationship is determined by the optical system's calibration matrix, called $\tilde{\mathbf{P}}$. Being *x*, *y*, *z* the coordinates to a point *P* in relation to an arbitrary system, and *i* and *j* the indexes to the corresponding *pixel P'* in image (Grewe and Kak, 1994):

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

where $j = U/S$ and $i = V/S$.

The calibration (or perspective transform) matrix is a function of some parameters which depend exclusively on the camera and of others related to the external system of coordinates. The first set, called *intrinsic parameters*, includes the distance from the optical center to the retina plane, *f*, the coordinates of

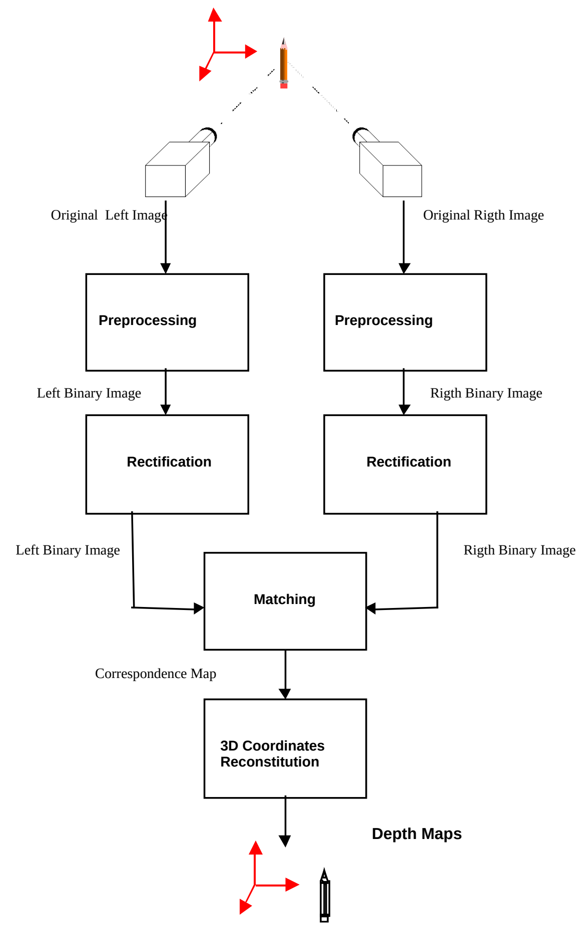

Fig. 3. Block diagram of the stereo vision system.

the main point (i_0, j_0), and the distances δu and δv from a pixel to its neighbors in the horizontal and in the vertical directions, respectively. The other set, called *extrinsic parameters*, includes the position C of the optical center with respect to the external system and the unitary vectors $\hat{\mathbf{h}}$, $\hat{\mathbf{v}}$ and $\hat{\mathbf{a}}$ which describe the camera orientation (figure 4).

In function of these variables, $\tilde{\mathbf{P}}$ is given by:

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

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

The reconstitution of a scene from two stereo images demands the perspective transformation matrixes $\tilde{\mathbf{P}}_1$ and $\tilde{\mathbf{P}}_2$ to be determined. Some variables the calibration matrix depends on cannot, however, be directly measured (Grewe and Kak, 1994). Although δu and δv are usually supplied by the camera manufacturer, only the two limits between which the distance f varies are furnished; its exact value is not. The position of the optical center inside the camera is also unknown, making it unfeasible to measure vector

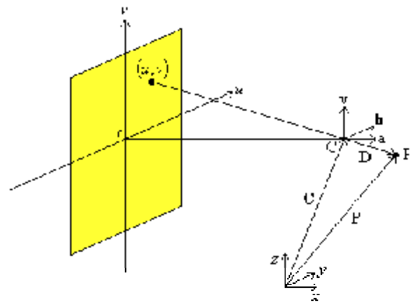

Fig. 4. Vectors that define the position and orientation of the camera with respect to the external coordinates system

C. Difficulty in accurately determining orientation of a camera's optical axis with regard to external system makes it impossible to obtain $\hat{\mathbf{h}}$, $\hat{\mathbf{v}}$, $\hat{\mathbf{a}}$ directly.

Camera calibration in positions A and B in figure 3 was carried on using an image of N reference points whose positions in relation to the adopted system of coordinates are known; next, by a linear system solution the matrix that maps those points to the indexes (i,j) to the pixels in image was determined.

Being (x_m, y_m, z_m) the coordinates of the m-th point and (i_m, j_m) the indexes to the corresponding *pixel* in image, equation (2) can be written as:

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

This is a system with $2N$ equations, the unknowns being the 12 elements in $\tilde{\mathbf{P}}$. Although 6 points are enough to solve it, 32 white squares on 2 orthogonal dark planes (figure 5) were used for robustness.

3.4 Stereo Images Correspondence

There are three kinds of techniques commonly used in determining stereo images correspondence, in machine vision systems (Horn, 1986). The first, called *feature matching*, is based on extracting from each image, separately, some feature that 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 (Redert, et al., 1999).

In this paper, the technique based on edge detection was chosen. This approach has led to the best results so far (Horn, 1986), and, although it generates sparse depth maps, it allows computational effort to be reduced a lot since only points extracted in the first stage are analyzed when defining the matching pairs.

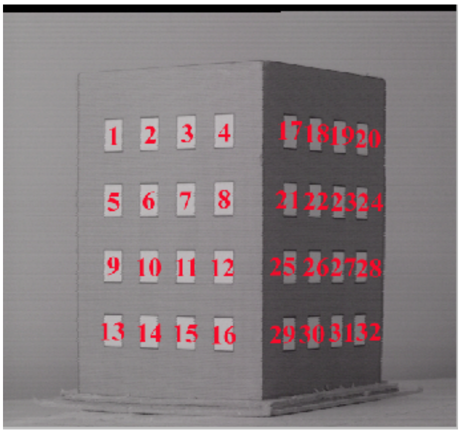

Fig. 5. Pattern with 32 reference points in 2 orthogonal planes, used in camera calibration.

Edge Detection in Stereo Images. Edge detection's aim is to determine the imaged object's frontiers by processing 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 (Iwahori, et al., 1999).

The procedure adopted to extract the stereo images' borders was to evaluate their derivatives in horizontal and vertical directions, using linear first-order filters. Regions where these derivatives were "high", according to fuzzy membership functions previously established, were generally interpreted as belonging to edges. Additional fuzzy rules were adopted in order to avoid double edges or isolated pixels in the output, resulting in greater robustness to input noise.

Rectification of Stereo Images. In this research, the algorithm used during the rectification process was the one proposed by (Fusiello, et al., 1998). 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 $\tilde{\mathbf{P}}_o = (\mathbf{P}_o|\mathbf{p}_o)$ and $\tilde{\mathbf{P}}_n = (\mathbf{P}_n|\mathbf{p}_n)$ are the perspective transformation matrices before and after rectification, respectively (Fusiello, et al., 1998):

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

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

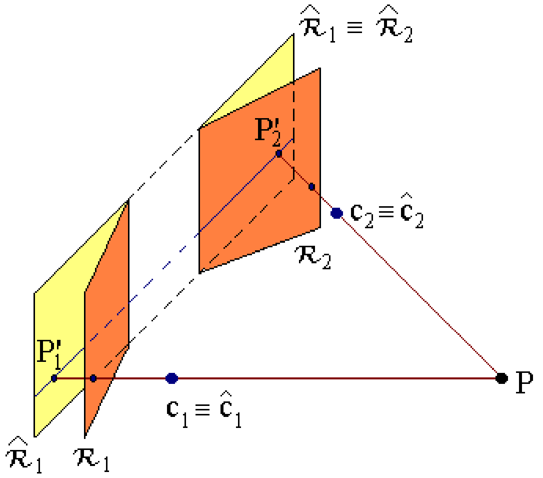

Fig. 6. Transformation of two stereo images' retina planes, in the rectification process.

By equation (5), the position of an object's image in each rectified retina plane is determined given its image in the plane in the original configuration.

Determining Homologue Pairs. The technique 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. Homologue pairs are then computed by sweeping all horizontal lines in the rectified images (figure 7). Since homologue points are always aligned, after rectification, an association of points in common lines in both images can be made (as figure 3 shows, polyhedrons' positions with respect to A e B guarantee that their edges appear in the same sequence in a horizontal line in the images).

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. The corresponding line was eliminated in this situation.

Notice that this algorithm has as inputs two binary images and as output a sparse matching map. But since the system was restricted to polyhedral forms, this map allowed the reconstitution not only of points in the edges but also of internal ones. The requisite was to compute lines that connect edge which are in the same horizontal line, in each rectified image. By this procedure, epipolar lines were reprojected in 3D space, allowing reconstituting points which do not belong to the matching maps.

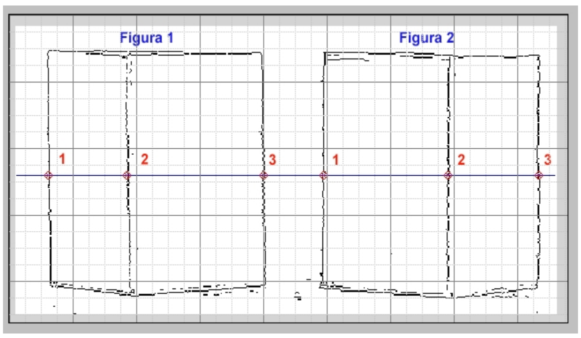

Fig. 7. Automatic matching of edges in the rectified stereo images.

3.5 Reconstitution of the Polyhedrons

After determining pairs of homologue points in the edges of the polyhedron's images, positions of the corresponding poi nts in the scene were computed from the intersection of the light rays from these to the optical centers (figure 2). This will now be stated algebraically: if (i_1, j_1) and (i_2, j_2) are homologue points in stereo i mages formed by the optical system $\mathbf{P_1}$–$\mathbf{P_2}$, the following is obtained from (2):

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

where $j_k = U_k / S_k$, $i_k = V_k / S_k$ ($k \in \{1,2\}$) and $[x\ y\ z]^T$ is the object with images (i_1, j_1) and (i_2, j_2).

If $\mathbf{w} = [x\ y\ z]^T$, $\mathbf{a}_n^T$ is the n th line in $\tilde{\mathbf{P}}_1$ and $\mathbf{b}_n^T$ is the n^th line in $\tilde{\mathbf{P}}_2$:

$$j_k = \frac{\mathbf{c}_1^T \cdot \mathbf{w} + c_{14}}{\mathbf{c}_3^T \cdot \mathbf{w} + c_{34}},\ i_k = \frac{\mathbf{c}_2^T \cdot \mathbf{w} + c_{24}}{\mathbf{c}_3^T \cdot \mathbf{w} + c_{34}}, \qquad (7)$$

where **c**=**a** if k=1 and **c**=**b** if k=2.

The solution to system in **w** (7), usually over - determined, leads to the coordinates $\mathbf{w} = [x\ y\ z]^T$.

4. RESULTS

The implemented stereo machine vision wa s applied to the reconstitution of the 3D -coordinates of a cube and of two prisms (with pentagonal and hexagonal bases, respectively). Figures 8 and 9 depict the stereo images, the detected edges and a 2D representation of *x*, *y*, *z* axes with all reconstituted points, for two cases (cube and prism with pentagonal basis). Results are compared to the known analyzed structures' models.

Since the most critical stage in the w hole 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 dete rmine when experimental errors were due to wrong matching or to imprecision in the calibration procedure or image acquisition.

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

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

Table 1 Distances obtained from the polyhedrons to the camera's optical centers in positions A and B (real value of 713mm for A and 714mm for 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 in calibration procedure (positions A and B).

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

5. 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 from the camera to the objects were estimated with errors under 3%, which is quite emough for robot navigation.

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

REFERENCES

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

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

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

Iwahori, Y., Md. S. Bhuiyan and A. Iwata (1999). Optimal edge detection under difficult imaging conditions. Technical report, Educ. Center for Inf. Processing and Dept. of Electrical and Computer Engineering. Nagoya Institute of Technology.

Redert, A., E. Hendriks and J. Biemond (1999). Correspondence estimation in image pairs — designing estimators for dense geometric correspondence fields. *IEEE Signa l Processing Magazine — 3D and Stereoscopic*, Visual Communication (Maio), **Vol. 16**, n° 3, pp 29–46.

APPENDIX — RESULTS OF THE RECONSTITUTION OF POLYHEDRONS

(a) Stereo Image, (b) Points reconstituted based on manual matching of stereo images (light gray), a nd known surface model (dark gray), (c) Detected borders, (d) Points reconstituted based on automatic matching (dark gray) and known surface model (light gray).

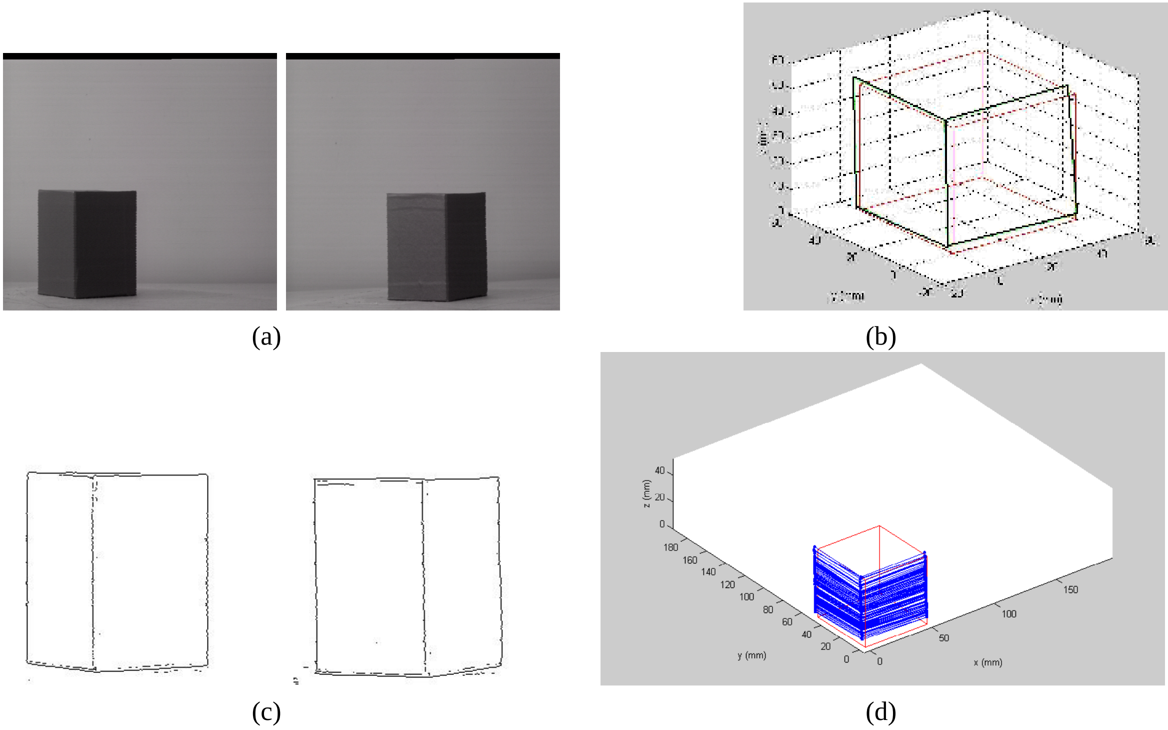

Fig. 8: Reconstitution of space coordinates of a cube, based on manual and automatic matching of a stereo image pair.

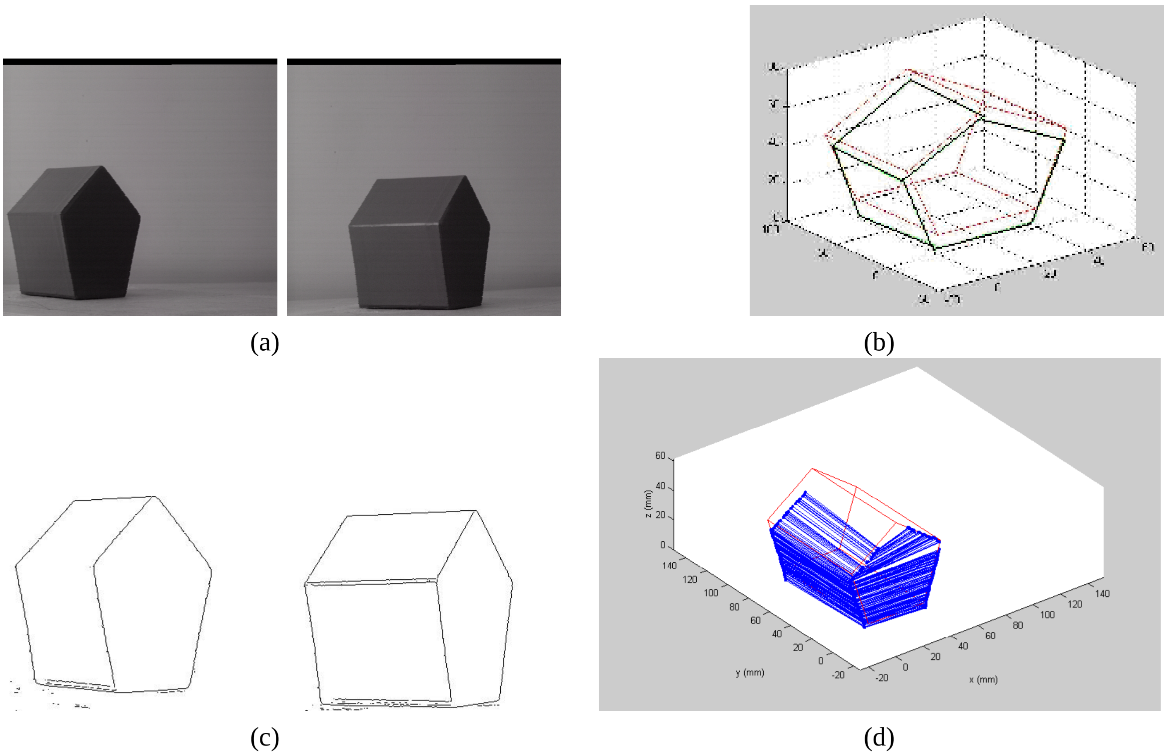

Fig. 9: Reconstitution of space coordinates of a pentagonal-base prism, based on manual and automatic matching of a stereo image pair.