Proceedings of the 2001 IEEE
International Conference on Robotics & Automation
Seoul, Korea - May 21-26, 2001

Uncalibrated Vision based on Structured Light

David Fofi*, Joaquim Salvi, El Mustapha Mouaddib***

*** CREA, Université de Picardie Jules Verne, Amiens (France)**
**** IIiA, Universitat de Girona, Girona (Spain)**
David.Fofi@sc.u-picardie.fr , qsalvi@eia.udg.es

Abstract

The recovery of the three-dimensional structure of the environment is a pre-requisite for many tasks in mobile robotics. Unfortunately, calibration acts as a brake upon visual adaptation and robotical autonomy. In this paper, we provide tools and constraints based on structured light for a complete and efficient uncalibrated Euclidean reconstruction of the environment. Experimental results achieved both on simulated and real data validate the method.

1 Introduction

Calibration was for many years the only way to recover three-dimensional scene structure from images. However, this approach suffers from two major drawbacks: firstly, the calibration process is very sensitive to errors and unstable; secondly, in many robotical applications, it is just not possible to calibrate on-line (for instance, if a calibration pattern is not available or if the camera is involved in visual tasks). Moreover, this process has to be repeated each time that one of the parameters is modified, which prohibits its use for visual adaptation, that is, working with a camera with automatic focus and aperture. As a consequence, reconstruction methods by means of uncalibrated sensors have been developped from the end of the eighties. For instance, Koenderink and Van Doorn [6] showed that the affine structure of a scene could be recovered from two images by computing an affine shape invariant. Faugeras [4] developed a reconstruction method from the knowledge of epipolar geometry. Mohr, Brand and Boufama [8] proposed a method of simultaneous retrieval of 3-D co-ordinates and projection matrices. Luong and Vieville [7] used a canonical representation of views based on the fundamental matrix in order to recover the projective structure of the scene.

Without any knowledge of scene geometry nor sensor parameters *but* pixel correspondences, it is demonstrated that it is possible to reconstruct three-dimensional scenes only up to a collineation; such a reconstruction is called *projective*. In order to upgrade a projective reconstruction into an Euclidean one, constraints on camera motion, intrinsic parameters or Euclidean scene geometry have to be added. The projection of a known regular pattern onto the scene permits to easily retrieve geometrical constraints which relate the object points. Moreover, it permits to artificially structure the observed surfaces. In this paper, we show how structured lighting is used to recover Euclidean structure of the observed scene from an uncalibrated sensor. As our purpose is to provide a practical and implementable method of reconstruction, we also give a set of useful tools for uncalibrated vision. The paper is structured as follows. In the section two, the structured light sensor is described. Next, tools for uncalibrated vision are presented. Then, the generation of Euclidean constraints by the analysis of coded images is detailed and experimental results are shown. Finally, the paper ends with conclusions.

2 Structured Light Vision

It is well known that the major drawback of stereoscopy is the correspondence problem, *i.e.* the matching of homologue points among the images. With the aim of reducing this drawback, coded structured light techniques have been developed [2]. In a structured light system, the second camera is replaced by a light source that projects a known pattern of light onto the scene, as shown in Figure 1. Since a projector can be seen as a camera *acting in reverse*, it can be modelled in the same way a camera is.

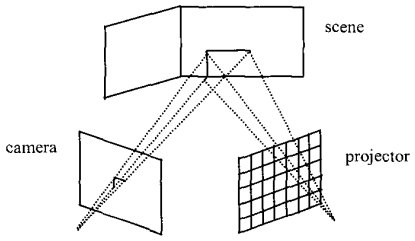


Figure 1. Structured Light Geometry

The pattern is composed by a set of vertical and horizontal slits, uniquely colour-encoded in a single pattern projection. The reader is pointed to Salvi *et al.* [9] to get deeper into the pattern design.

3 Tools for Uncalibrated Vision

This section first recalls the definition of the cross-ratio and its basic formulations. Then, it is described how to use it to perform a test of spatial colinearity and a test of coplanarity only from pixel co-ordinates.

3.1 The Cross-Ratio

The cross-ratio is the fundamental projective invariant. It is a numerical value, computed from a certain configuration of points, which remains unchanged through homographies (Figure 2). The cross-ratio $k = \{P; A, B, C, D\}$ of a pencil of four convergent lines in a plane is defined by:

$$k = \frac{\sin(\hat{APC})}{\sin(\hat{BPC})} \cdot \frac{\sin(\hat{BPD})}{\sin(\hat{APD})} \quad (1)$$

With a line crossing the pencil and A', B', C', D', the subsequent intersections, the same cross-ratio can be computed by:

$$k = \{A', B', C', D'\} = \frac{\overline{A'C'}}{\overline{B'C'}} \cdot \frac{\overline{B'D'}}{\overline{A'D'}} \quad (2)$$

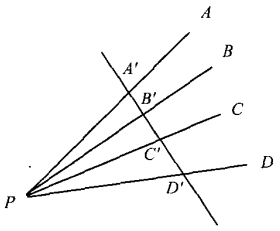


Figure 2. The Cross-Ratio

The cross-ratio can be used for obtaining two geometrical constraints : a test of spatial colinearity and a test of coplanarity. They are both described in the following of this section; an analysis of the stability of the cross-ratio with respect to these tests is also proposed.

3.2 Colinearity and Coplanarity

Let us consider four points in space P, Q, R, S, their respective corresponding points p, q, r, s in the pattern projector frame and their projections p', q', r', s' onto the image plane (Figure 3). Whether P, Q, R and S are aligned, it is obtained:

$$k = \{P, Q, R, S\} = \{p, q, r, s\} = \{p', q', r', s'\} \quad (3)$$

The change from points of the pattern projector frame to points of 3-D space is achieved through an homography; likewise, the change from points of 3-D space to points of the image plane is also achieved through an homography. As cross-ratios are preserved through homographies, it can be inferred that the three values are equal. Conversely, it can be inferred that whether the cross-ratio of four imaged points is equal to the cross-ratio of the four corresponding projected points, the points P, Q, R and S belong to the same straight line in space. Since the co-ordinates of the projected points are perfectly known, structured light provides an accurate measure of this cross-ratio. The test of colinearity will simply consist in checking the value of the cross-ratio of four points within the image.

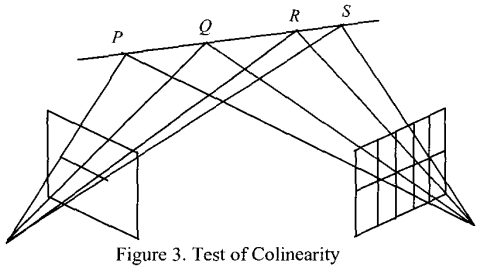


Figure 3. Test of Colinearity

Now, let us consider the two configurations of points shown in Figure 4. Whether the points o_i, p_i, q_i, r_i et s_i (i = 1 or 2) are projected onto a plane, the cross-ratio within the pattern is equal to the cross-ratio of the five points formed onto this plane; moreover, the cross-ratio of the homologue points within the image is equal to both. Once more, the change from projected points to imaged points is obtained by two successive homographies.

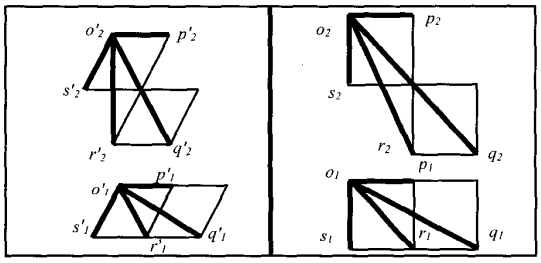


Figure 4. Test of Coplanarity

It can be deduced that if equation (4) is verified, the corresponding object points O, P, Q, R and S are coplanar.

$$\{o_i; p_i, q_i, r_i, s_i\} = \{o'_i; p'_i, q'_i, r'_i, s'_i\} \text{ with } i = 1 \text{ or } 2 \quad (4)$$

We have tested the stability of the cross-ratio for both configurations of points, *i.e* for the test of colinearity and for the test of coplanarity. For the first, we took four aligned points for which the distances d between two successive points are equal. A noise, varying from 0 to 0.5 × d, is added on the points co-ordinates. The results are illustrated by Figure 5 (Top). For the second, experimental protocol remains unchanged: only the points configuration changes and, consequently, the formula used for the cross-ratio. The results are gathered in Figure 5 (Bottom).

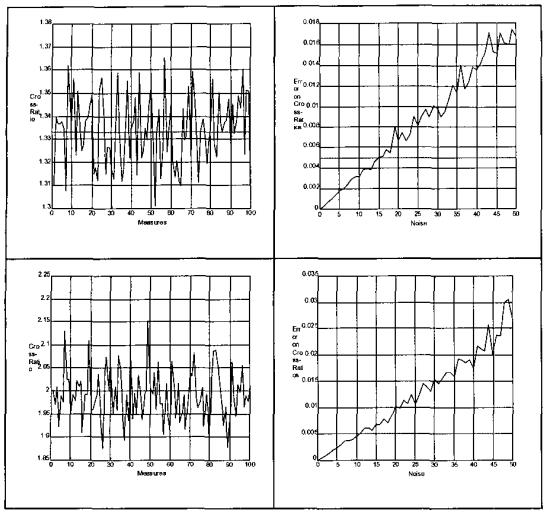


Figure 5. (Top) Test of colinearity (Bottom) Test of coplanarity (Left) Measures with error ± 5% (Right) Measuring error with a noise from 0 to 50%.

To be able to compare the theoretical cross-ratios with the cross-ratios computed from the images, we used a *projective distance* based on the method of the random cross-ratios [1]. The tolerance error is empirically fixed to 5×10^{-3}. Under these conditions, a noise up to ±15% is allowed to well discriminate configurations of colinear and coplanar points. Obviously, as it can be deduced from the results of Figure 5, the more the distance d is large, the more the measure of cross-ratio is robust. The left part of Figure 5 shows that, with a moderate noise (± 5%), the measured cross-ratio is very near to the theoretical one. Hence, the stability of cross-ratio is good enough for applications in uncalibrated reconstruction.

4 System Modelling

This section presents the affine camera model and proposes a test of validity for this model. The purpose is to provide a method to verify the affine assumption without any knowledge of scene geometry nor sensor parameters.

4.1 The Affine Model

In the affine camera model, the projection is not performed through a fixed point but along a direction. In fact, it corresponds to the case when the focal point is set at infinity. This is equivalent to say that the model does not take into account the depth of the scene. Whether the direction is orthogonal to the image plane, the projection is said orthographic. It leads to a very simple formulation:

$$\begin{cases} u = X \\ v = Y \end{cases} \tag{5}$$

where (u, v) are pixel co-ordinates and (X, Y, Z) are object point co-ordinates. This model preserves parallelism and relative distances.

Fixing the focal point at infinity imposes an important constraint: the depth ratio, seen as the ratio between the length of the scene along the optical axis and the scene-camera distance has to be small enough (more or less 1/20). The next sub-section presents a method to test the validity of the affine model from image analysis.

4.2 Test of Validity

This section describes a method to compute a confidence rating for the validity of the affine model. An affine transformation change a square into a parallelogram, in the worst case. Hence, a square projected onto a planar surface and captured by a camera will form a parellelogram onto the image plane whether the sensor follows the affine model.

Let us consider the four vertices of a square of the pattern and let us assume that they are projected onto a planar surface. Their corresponding points within the image plane form a quadrilateral. If it appears to be a perfect parallelogram, the affine model can be assumed; otherwise, the hypothesis has to be discussed or gave up. What about intermediate forms? And how to correlate the square deformation with the validity of the affine model?

In concrete terms, let us imagine a parallelogram whose vertices are the points m, n, n' and m', succesively and clockwise. Lengths of opposite edges of a parallelogram are equal. By measuring the difference between these edges length with equation (6), knowledge of parallelogram-like forms can be retrieved:

$$\phi = \frac{\left|\sqrt{(u_m-u_n)^2+(v_m-v_n)^2}-\sqrt{(u_{m'}-u_{n'})^2+(v_{m'}-v_{n'})^2}\right|}{\sqrt{(u_m-u_n)^2+(v_m-v_n)^2}} + \frac{\left|\sqrt{(u_m-u_{m'})^2+(v_n-v_{n'})^2}-\sqrt{(u_m-u_{m'})^2+(v_n-v_{n'})^2}\right|}{(u_m-u_{m'})^2+(v_n-v_{n'})^2} \tag{6}$$

where $(u_x, v_x)_{x=m,n,m',n'}$ are the pixel co-ordinates of each point. We assume that the affine model is valid if ϕ is near to zero.

5 Uncalibrated Reconstruction

Let us consider the co-ordinates vector $\tilde{\mathbf{M}}$ of a point projectively reconstructed (from a structured light sensor, it has been shown that the reconstruction method has to be performed from a single pattern projection and a single image capture [5]). Since Euclidean transformations form a sub-group of projective transformations, it is obtained:

$$\mathbf{M} = \mathbf{W} \cdot \tilde{\mathbf{M}} \tag{7}$$

where $\mathbf{M}$ is the vector of co-ordinates of the same point expressed in an Euclidean frame. $\mathbf{W}$ is a collineation,

i.e. an invertible 4×4 matrix defined up to a scale factor. The Euclidean constraints method [3] consist of translating geometrical knowledge of the 3-D points into mathematical constraints on the entries of **W**. In the following, we present a method which permits to recover constraints by the analysis of coded images. The sensor behaviour is assumed to be affine; the test of the fourth section can be used to verify the validity of this assumption. Moreover, the tools presented in section three can be used as well as Euclidean constraints or as means to choose projective basis (five points, no four of them being coplanar and no three of them being colinear) required by some projective reconstruction methods.

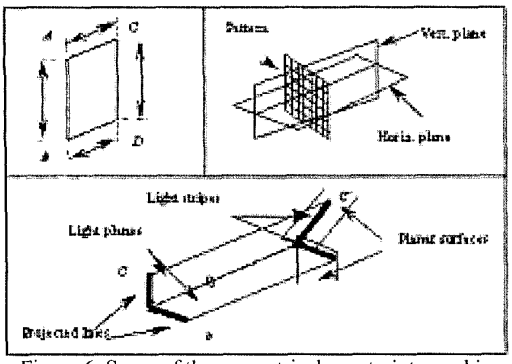


Figure 6. Some of the geometrical constraints used in Euclidean Reconstruction

5.1 Parallelogram Constraints

Projecting a square onto a planar surface, the more generic quadrilateral formed onto the surface is a parallelogram whether an affine model is assumed. Furthermore, a parallelogram captured by an affine camera forms a parallelogram onto the retina. Hence, a parallelogram within the image corresponds to the image of a parallelogram on a 3-D plane; the test of coplanarity is used to verify that the points are effectively coplanar.

Relative positioning of the four points A, B, C and D in space (Figure 6a) is such as:

$$\overline{AB} = \overline{CD},\ \overline{AC} = \overline{BD} \tag{8}$$

$$(AB)//(CD),\ (AC)//(BD) \tag{9}$$

It leads to a *redundant* set of constraints on **W**. As projective geometry keep unchanged alignment and coplanarity, equations (8) and (9) determine the same configuration of points. It has to be said that, likewise, a parallelogram completely determines a 3-D plane. Therefore, for each plane composing the scene, a unique set of parallelogram constraints is sufficient.

5.2 Orthogonality Constraint

Orthogonality is an important feature for Euclidean reconstruction. The detection of orthogonal planes permits to define, at least partially, a 3-D Euclidean frame of the scene. Let us consider again an affine model for the projector. The projection of a line produces a light plane in space. The projection of two orthogonal lines (AB) and (AC) produces two orthogonal light planes (Figure 6c).

When light planes intersect planar surfaces, they produce light stripes on them whose will be imaged by the camera. We have thus two lines $(A'B')$ and $(A'C')$ in space, which belong to orthogonal planes. Since A' and B' belong to the same horizontal plane and A' and C' belong to the same vertical plane, whether the world co-ordinate system is fixed at the projector, it is obtained:

$$x_{A'} = x_{B'} \quad ; \quad y_{A'} = y_{C'} \tag{10}$$

$$\overrightarrow{A'B'} \cdot \overrightarrow{A'C'} = (x_{A'} - x_{B'})(x_{A'} - x_{C'}) + (y_{A'} - y_{B'})(y_{A'} - y_{C'}) + (z_{A'} - z_{B'})(z_{A'} - z_{C'}) = (z_{A'} - z_{B'})(z_{A'} - z_{C'}) \tag{11}$$

Thus:

$$(A'B') \perp (A'C') \Leftrightarrow z_{A'} = z_{B'} \text{ or } z_{A'} = z_{C'} \tag{12}$$

If the conditions imposed by (12) are satisfied, we obtain an *orthogonality constraint*, otherwise we obtain the following *reduced orthogonality constraint*:

$$(x_A - x_B)(x_A - x_C) + (y_A - y_B)(y_A - y_C) = 0 \tag{13}$$

5.3 Horizontal Plane, Vertical Plane

Each projected horizontal or vertical line of the pattern generates a light plane in space, which can be considered as a horizontal or vertical 3-D plane in the projector co-ordinate system, respectively (Figure 6b). Indeed, what it is imaged by the camera are the intersections of light planes with the surfaces composing the scene, therefore points belonging to horizontal and/or vertical planes. If a point P belongs to the horizontal plane of the Euclidean frame in which the scene will be reconstructed, then $y_A = 0$. Likewise, the homologue constraint $x_A = 0$ can be expressed. Besides, two points belonging to the same plane have a component in common, which provides a constraint between the co-ordinates of both points.

Furthermore, an arbitrary distance can be set between two successive horizontal planes or vertical planes. The distance between two points A and B is given by:

$$(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 = d^2 \tag{14}$$

Finally, the cross-point (which appears in the image as the intersection of two light stripes) of the planes $y = 0$ and $x = 0$ can be defined as the origin of the Euclidean frame by equalling its three components to zero.

6 Experimental Results

6.1 Colinearity and Coplanarity

The results were obtained from images captured under realistic conditions (the coded images shown throughout the paper are negative ones). Beforehand, the test of spatial colinearity is evaluated. On Figure 7, three different configurations of points are tested. The first one (from up to bottom) represents four aligned points on a plane: the measured cross-ratio is very close to the theoretical one and the colinearity is detected. In the second example, where the pattern is projected onto an edge of a cube, the points are classified as not colinear. In the third example, where the points are clearly not colinear, the projective error even reach infinity.

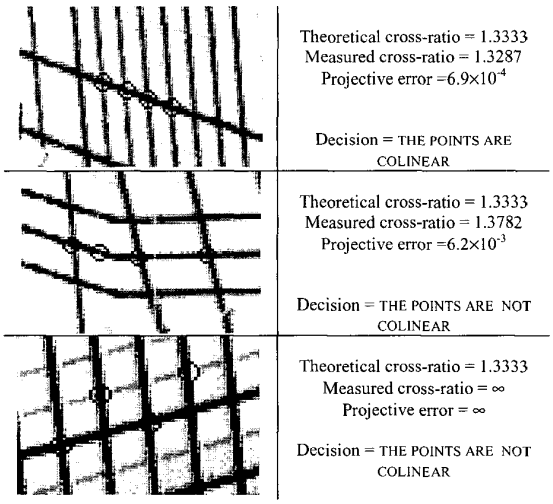


Figure 7. Results of the test of colinearity

Moreover, Figure 8 shows three examples of coplanarity test. In the first one, a plane was lighted by the pattern: the test detects a planar configuration of points. In the second example, it is noted that the pattern is projected onto an irregular surface; and the points are classified as non-coplanar. Finally, the third example represents an edge of a cube where the projective error is important.

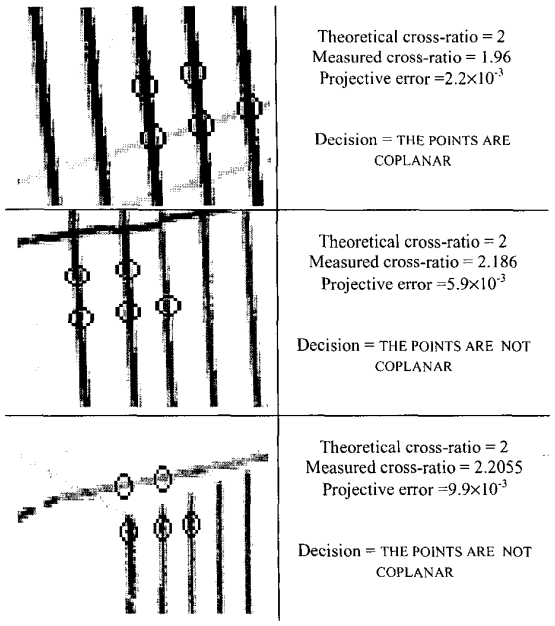


Figure 8. Results of the test of coplanarity

6.2 Reconstruction

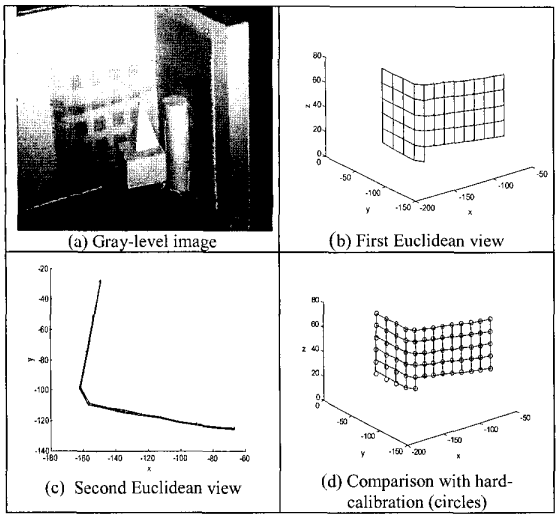
Firstly, we have achieved the reconstruction method with simulated data, using the parameters estimation approach of Mohr *et al.* [8] to perform the projective reconstruction and adding Euclidean constraints from structured lighting. The table 1 shows the results we obtained for ten points of the scene: mean relative error is less than 2% with a uniform noise added on pixel co-ordinates.

Real co-ordinates			Errors on estimate co-ordinates		
X	*Y*	*Z*	*ΔX*	*ΔY*	*ΔZ*
100	-50	4000	0.518	-0.267	3.95
300	-50	2000	-0.65	-0.242	-1.5
700	-50	4000	0.614	-0.33	6.43
500	-400	4020	-1.132	-1.768	-4.332
300	50	4000	0.091	0.397	2.597
500	50	2000	0.076	-0.119	0.449
900	50	4000	0.13	0.171	2.007
300	-430	3000	0.505	-0.911	5.079
450	75	2500	0.76	-1.154	4.016
705	-120	1000	0.603	-0.827	0.829
Mean relative error (%)			**0.518**	**1.539**	**0.169**

Table 1: Errors on reconstruction with uniform noise ± 1

Secondly, we have reconstructed 70 points of a real scene, which represents a cube (see Figure 9a). We have performed the Euclidean reconstruction using the constraints defined in the previous section. To illustrate the accuracy of the method, a comparison between hard-calibrated and uncalibrated reconstruction is presented on Figure 9d. As we did not have real 3-D co-ordinates but only 3-D co-ordinates computed by hard-calibration, a quantitative evaluation of the reconstruction method is not possible. However, the comparison with results

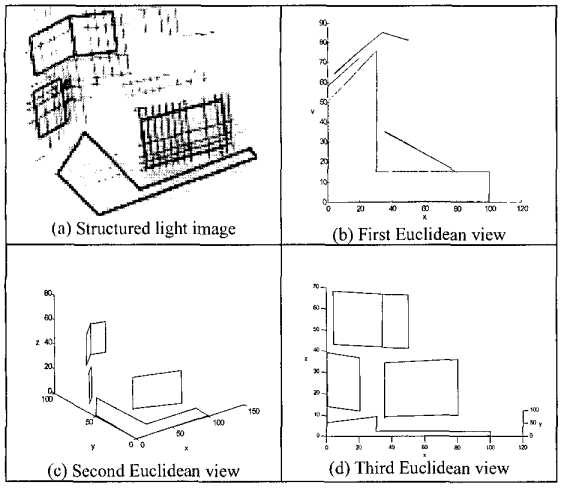
obtained by hard-calibration shows that uncalibrated reconstruction from grid coding obtains good results (in a qualitative way). On Figure 10, we present results of Euclidean reconstruction performed from an image grabbed under realistic condition. Only the highlighted lines have been reconstructed. It can be noted that parallelism and orthogonality are well-recovered and relative distances are respected. Only a few iterations are necessary to perform a Euclidean reconstruction from a projective one.



(a) Gray-level image (b) First Euclidean view

(c) Second Euclidean view (d) Comparison with hard-calibration (circles)

Figure 9. Euclidean Reconstruction by geometrical constraints



(a) Structured light image (b) First Euclidean view

(c) Second Euclidean view (d) Third Euclidean view

Figure 10. Euclidean Reconstruction by geometrical constraints

7 Conclusions

This paper presents a complete method to perform Euclidean reconstruction from an uncalibrated structured light sensor. It describes a method to obtain such a reconstruction by only analysing the coded images.

It is known that a collineation exists which upgrades projective reconstruction to Euclidean one. This collineation can be assessed by translating geometrical information about the scene into constraints on the entries of the collineation matrix. We show that geometrical knowledge of the scene as parallelism or orthogonality can be retrieved by projecting a known pattern of light onto the scene. Moreover, the paper provides tools to make the method more efficient and automatic. By this way and assuming that the images are well-segmented, it is possible to reconstruct the scene without any human intervention.

As no constraints are required on projection matrices, the presented approach allows us to reconstruct changing the focus, the aperture and the zoom of both the camera and the projector. Therefore, the sensor can be involved in visual tasks which require self-adaptability with numerous applications as autonomous navigation, visual exploration, among others.

8 References

[1] K. Aström, L. Morin, "Random cross-ratios", Research Report n°RT 88 IMAG-14, LIFIA, 1992.

[2] J. Batlle, E. Mouaddib, J. Salvi, "Recent progress in coded structured light to solve the correspondence problem. A survey", Pattern Recognition, 31(7), pp. 963-982, 1998.

[3] B. Boufama, R. Mohr, F. Veillon, "Euclidean constraints for uncalibrated reconstruction", Proc. of the 4th Int. Conf. on Computer Vision, Berlin (Germany), pp. 466-470, 1993.

[4] O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?", Proc. of the 2nd Euro. Conf. on Computer Vision, Santa Maria Ligure (Italia), pp. 563-578, 1992.

[5] D. Fofi, J. Salvi, E. Mouaddib, "Euclidean reconstruction by means of an uncalibrated structured light sensor", Proc. of the Vth Ibero-American Symposium on Pattern Recognition, Lisboa, 2000.

[6] J.J Koenderink, A. J. Van Doorn, "Affine structure from motion", Technical Report, Utrecht University, Utrecht (Netherlands), october 1989.

[7] Q.-T. Luong, T. Viéville, "Canonical representations for the geometries of multiple projective views", Proc. of the 3rd Euro. Conf. on Computer Vision, Stockholm (Sweden), 1994.

[8] R. Mohr, B. Boufama, P. Brand, "Accurate projective reconstruction", Proc. of the 2nd ESPRIT-ARPA-NSF Workshop on Invariance, Azores, pp. 257-276, 1993.

[9] J. Salvi, J. Batlle, E. Mouaddib, "A robust-coded pattern projection for dynamic measurement of moving scenes", Pattern Recognition Letters, 19, pp. 1055-1065, 1998.