EUCLIDEAN RECONSTRUCTION BY MEANS OF AN UNCALIBRATED STRUCTURED LIGHT SENSOR

David Fofi[1], Joaquim Salvi[2], El Mustapha Mouaddib[1]
[1] Perception Group, CREA, Université de Picardie Jules Verne,
7 rue du Moulin Neuf, 80000 Amiens (France)
[2] Computer Vision and Robotics Group, IIiA, Universitat de Girona,
Avda. Lluis Santalo, s/n, 17071 Girona (Spain)
{David.Fofi, mouaddib}@sc.u-picardie.fr, qsalvi@eia.udg.es

Abstract

A method of uncalibrated reconstruction through structured lighting is presented in this paper. The co-ordinates of 3-D points and the projection matrices are simultaneously determined in a parameter estimation approach. The reconstruction is performed in a projective frame, up to a projective transformation, by only using pixel correspondences. It is shown that an Euclidean reconstruction can be recovered by constraining the projective transformation by using geometrical knowledge about the scene. Moreover, it is described how the pattern projection is used to grab this kind of knowledge. This method provides our sensor with adaptive capabilities and permits to be used in mobile robotics. Experimental results realised both on simulated and real data are presented.

1. INTRODUCTION

The perception of the three-dimensional structure of the environment is an important task in computer vision. In mobile robotics, it forms the basis of obstacle detection, map building, scene analysis, etc. The drawback is to infer 3-D information of a scene by starting from two images of it, at least. A solution consists in calibrating the camera, *i.e.* determining its optical parameters and internal geometry (focal distance, principal point, pixel adjustment) and its geometrical parameters (position and orientation with respect to a reference frame). With this goal, the correspondence among the 3D object points and their projections has to be established; the obtained matrix makes possible to connect each point to its line of sight [5]. From two images, the reconstruction is thus possible. This technique is called *hard-calibration* and is carried out off-line by using a calibrating pattern whose 3-D points co-ordinates of interest are precisely known. This process has to be repeated each time that one of the parameters of the camera is modified. Hard-calibration is exclusively adapted to applications that keep the sensor unchanged during the measuring process. Nevertheless, a visual adaptation to the environment is essential in mobile robotics. The visual adaptation permits to work with a camera with auto-focus (to increase the quality of the image), zoom (to concentrate on

relevant regions of the image) and aperture (in case of illumination changes), which is the first step to develop strategies of observation and/or exploration.

With the aim of avoiding calibration process, reconstruction techniques using uncalibrated sensors increased in importance since the end of the eighties. They permit to perform reconstruction on-line, and thus to be used in mobile robotics and adaptive vision. Pixel correspondences and some geometrical knowledge about the sensor and/or the scene are sufficient to achieve this task. In 1989, Koenderink and Van Doorn [7] showed that the affine structure of a scene could be computed from two images, assuming an affine camera model. Many methods based on affine camera model were developed since then [10,11,12]. In 1992, Faugeras [3] developed a reconstruction method adapted to the pinhole model using tools of projective geometry. The epipolar geometry of the stereo rig is assumed known. The author showed that this kind of reconstruction could only be performed up to a projective transformation. The same year, Hartley et al. [6] reached similar results. In 1993, Mohr *et al.* [8] proposed a method of simultaneous retrieval of 3-D co-ordinates and matrix parameters. Boufama et al. [2] showed that an Euclidean reconstruction can be performed through the knowledge of Euclidean characteristics of the scene starting from a projective reconstruction.

The main goal of this work is, firstly, to provide adaptive capabilities to our vision sensor and secondly to adapt the techniques of uncalibrated reconstruction to the structured light sensors. The major contribution of the article is to demonstrate how the projection of a grid of light and the analysis of the coded images permit to generate Euclidean constraints for a three-dimensional reconstruction. This article is organized as follows. The section two presents the structured light system. Then, section three details the uncalibrated reconstruction method. Furthermore, section four deals with Euclidean reconstruction through structured lighting. Finally, section five presents experimental results. The article ends with conclusions.

2. STRUCTURED LIGHT SYSTEM

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 problem, coded structured light techniques have been developed [1]. 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 1a. Since a projector can be seen as a camera acting in reverse, it can be modelled in the same way a camera does. The pattern is composed by a set of vertical and horizontal slits, uniquely colour-encoded in a single pattern projection (figure 1b). The reader is pointed to Salvi *et al.* [9] to get deeper into the pattern structure. The pattern codification permits to solve the correspondence problem knowing for each imaged point, its corresponding point on the emitted projector plane, by decoding the pattern. It has to be noted that any movement of the projector during the calibrating process will produce a movement of the pattern projected onto the scene. As the

correspondence is obtained from the light projected, a projector movement will produce a movement of the 3-D object points: thus, structured light vision imposes that the reconstruction has to be performed from only one view and a single pattern projection (*i.e.* two views in stereovision).

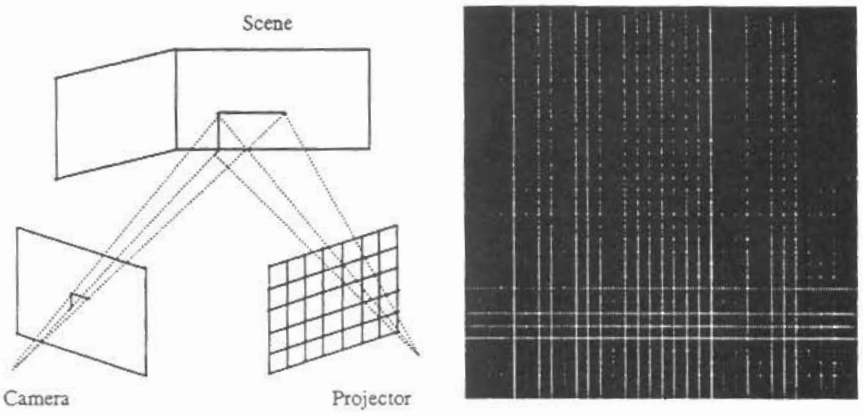

Figure 1: The structured light system: (a) the geometrical principle; (b) the colour-encoded pattern.

3. PROJECTIVE RECONSTRUCTION

Let us consider n images ($n \geq 2$) composed by m points. More precisely, it is sufficient that each point appears in two images. The pixel co-ordinates of these points are supposed to be matched among the images. It is obtained that:

$$\mathbf{p}_{ij} \approx \mathbf{A}_j \cdot \mathbf{P}_i \,,\; i \in \{1,\ldots,m\} \; j \in \{1,\ldots,n\} \tag{1}$$

The symbol $\approx$ denotes an equality up to a scale factor also called *projective equality*, $\mathbf{p}_{ij} = \begin{pmatrix} u_{ij} & v_{ij} & w_{ij} \end{pmatrix}^T$ and $\mathbf{P}_i = \begin{pmatrix} x_i & y_i & z_i & t_i \end{pmatrix}^T$. $\mathbf{A}_j$ is the projective matrix associated with the j^{th} image:

$$\mathbf{A}_j = \begin{bmatrix} a_{11}^{(j)} & a_{12}^{(j)} & a_{13}^{(j)} & a_{14}^{(j)} \\ a_{21}^{(j)} & a_{22}^{(j)} & a_{23}^{(j)} & a_{24}^{(j)} \\ a_{31}^{(j)} & a_{32}^{(j)} & a_{33}^{(j)} & a_{34}^{(j)} \end{bmatrix} \tag{2}$$

Developing this equation and re-arranging it in order to eliminate the homogeneous co-ordinates, $U_{ij} = u_{ij} / w_{ij}$ and $V_{ij} = v_{ij} / w_{ij}$ are obtained:

$$\begin{cases} U_{ij} = \dfrac{a_{11}^{(j)} x_i + a_{12}^{(j)} y_i + a_{13}^{(j)} z_i + a_{14}^{(j)} t_i}{a_{31}^{(j)} x_i + a_{32}^{(j)} y_i + a_{33}^{(j)} z_i + a_{34}^{(j)} t_i} \\ V_{ij} = \dfrac{a_{21}^{(j)} x_i + a_{22}^{(j)} y_i + a_{23}^{(j)} z_i + a_{24}^{(j)} t_i}{a_{31}^{(j)} x_i + a_{32}^{(j)} y_i + a_{33}^{(j)} z_i + a_{34}^{(j)} t_i} \end{cases} \quad (3)$$

where (U_{ij}, V_{ij}) are the co-ordinates of the i^{th} point in the j^{th} image.

As we have m space points and n images, the number of equations is $2 \times m \times n$. The unknowns are $11 \times n$ for the projection matrices (defined up to a scale factor, so $3 \times 4 - 1$ degrees of freedom) and $3 \times m$ for the points to be reconstructed. By choosing m and n large enough, we can reach a redundant set of equations. Note that if $\mathbf{A} \cdot \mathbf{P}$ is a solution, then $\mathbf{AW}^{-1} \cdot \mathbf{WP}$ is also a solution:

$$\mathbf{p} = \mathbf{A} \cdot \mathbf{P} \Leftrightarrow \mathbf{p} = \left(\mathbf{AW}^{-1}\right) \cdot \left(\mathbf{WP}\right) \quad (4)$$

$\mathbf{W}$ is a 4×4 invertible matrix which represents a collineation of the 3-D space. Therefore, this kind of reconstruction can only be performed *up to a collineation*. Hence, in order to perform a *unique* solution, $\mathbf{W}$ has to be determined. Since $\mathbf{W}$ has 15 degrees of freedom, five points in space are sufficient to determine it. With no *a priori* knowledge about the scene, five 3-D points (no four of them being coplanar) are taken and their co-ordinates are assigned to an arbitrary projective basis. In these conditions, the reconstruction is achieved in a projective frame. If the Euclidean co-ordinates of five points are known, we can directly obtain a reconstruction in an Euclidean frame, also called *Euclidean reconstruction*. It is known that the Euclidean geometry is a particular case of the projective geometry, in other words, a collineation $\mathbf{W}$ exists which brings the solution to an Euclidean one.

$$\mathbf{P} = \mathbf{W} \cdot \mathbf{P}' \quad (5)$$

where $\mathbf{P}'$ denotes a 3-D point reconstructed in a projective frame and $\mathbf{P}$ the same point reconstructed in an Euclidean frame. So as to compute the collineation, geometrical knowledge about the scene is translated into constraints on the entries of $\mathbf{W}$ (see [2] for further information).

4. EUCLIDEAN CONSTRAINTS THROUGH STRUCTURED LIGHTING

The interest of this section is to demonstrate that geometrical knowledge about the scene can be obtained analysing the grid coded images. This knowledge can be used as constraints for an Euclidean reconstruction.

4.1. Plane detection in space

The elementary cell of a grid is a square. Each cell projected onto a planar surface is captured by the camera as a quadrilateral. Quadrilaterals can theoretically either be captured as squares, rectangles, rhombuses, parallelograms or trapezoids, depending on the position and the orientation of the camera, the projector and the planar surface observed. We assume that quadrilateral detection within the image is equivalent to plane detection in 3-D space although there are configurations of curved surfaces that can yield quadrilaterals in the image, and where also the four vertices are not coplanar even. According to the planarity constraint [4], it is possible to retrieve plane orientation through the knowledge of R and t (*i.e.* the orientation and translation from the camera to the projector).

4.2. Parallelogram constraint

Assuming that the projector have an approximately affine behaviour, we obtain that projecting a square onto a planar surface, the more generic quadrilateral formed onto this plane is a parallelogram. 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. Geometrical knowledge about the scene can thus be deduced.

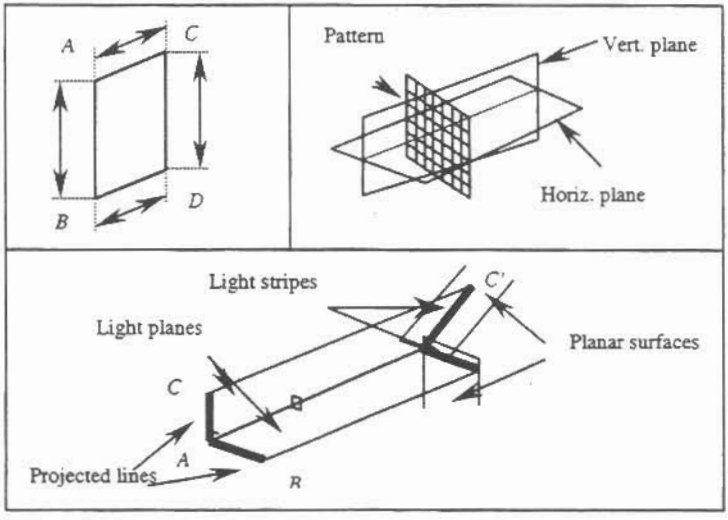

Figure 2: Some of the geometrical constraints used in Euclidean reconstruction: (a) Parallelogram constraint (b) Horizontal and vertical plane (c) Orthogonality

Relative positioning of the four points A, B, C and D (see figure 2a) in space is such as:

$$\overline{AB} = \overline{CD},\ \overline{AC} = \overline{BD} \tag{6}$$

$$(AB)//(CD),\ (AC)//(BD) \tag{7}$$

It leads to a *redundant* set of constraints on $\mathbf{W}$. As projective geometry keep unchanged alignment and coplanarity, equation (6) and (7) 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.

4.3. Horizontal and vertical plane

If a point P belongs to the horizontal plane passing through the origin, then $z_P = 0$, which permits to obtain another linear constraint. The homologue constraints for $y_P = 0$ and $x_P = 0$ can obviously be expressed. Noting that each projected horizontal line of the pattern generates a light plane in space, which can be considered as a horizontal 3-D plane in the projector co-ordinate system (see figure 2b). And each projected vertical line of the pattern generates a light plane in space, which can be considered as a vertical 3-D plane in the projector co-ordinate system (considering either an affine or projective camera model). Their corresponding lines captured by the camera can be used to generate such kind of constraints. Indeed, what it is imaged by the camera are the intersections of light planes generated by the pattern projection with the surfaces composing the scene, therefore points belonging to horizontal and/or vertical plane.

Furthermore, an arbitrary distance can be set between two successive horizontal planes or vertical planes. If the distance between two points A and B is d, we find the following non-linear constraint:

$$(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 = d^2 \tag{8}$$

This constraint permits to assign a metric to 3-D space. It is possible to give an arbitrary value to d but the reconstruction will be achieved up to a scale factor.

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.

4.4. Orthogonality constraint

Orthogonality is an important feature for Euclidean reconstruction. Detecting orthogonal planes permits to define, at least partially, 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 2c). 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:

$$x_{A'} = x_{B'} \quad ; \quad y_{A'} = y_{C'} \tag{9}$$

$$\overrightarrow{AB}\cdot\overrightarrow{AC} = (x_A - x_B)(x_A - x_C) + (y_A - y_B)(y_A - y_C) + (z_A - z_B)(z_A - z_C) = (z_{A'} - z_B)(z_A - z_C) \tag{10}$$

So:

$$(A'B') \perp (A'C') \Leftrightarrow z_{A'} = z_{B'} \text{ or } z_{A'} = z_{C'} \tag{11}$$

If the conditions imposed by (11) are satisfied, we obtain an *orthogonality constraint*, otherwise we obtain a *reduced orthogonality constraint*:

$$(x_{A'} - x_{B'})(x_{A'} - x_{C'}) + (y_{A'} - y_{B'})(y_{A'} - y_{C'}) = 0 \tag{12}$$

5. EXPERIMENTAL RESULTS

5.1. Simulated data

The reconstruction is performed in two steps: a projective reconstruction assigning the reference points to an arbitrary projective basis and then an Euclidean reconstruction performed from the previous projective reconstruction to which geometrical knowledge about the scene is added. The Levenberg-Marcquardt algorithm is used to solve the set of non-linear equations (see [2] for detailed explanations). Projective reconstruction provides projection matrices and 3-D co-ordinates with respect to a projective frame. In order to validate this reconstruction, the 3-D co-ordinates are retroprojected onto the image planes through the projection matrices and the residual error is evaluated (see figure 3). Our conclusion is that projective reconstruction performed well in most cases. However, in order to ensure convergence of the algorithm, the relative positioning of the 3-D points must correspond more or less to the configuration of the chosen basis *i.e.* the Euclidean reference points configuration must be in adequacy with the projective co-ordinates given to them. The 3-D point co-ordinates are initialised as the co-ordinates of the barycenter of points that must be reconstructed. Obviously, with simulated data and no noise, the discrepancy between the real 3-D points and the reconstructed 3-D points are nearly zero. The error is due to round off in digital computation. So, in

order to analyse the robustness of the method, noise is added on pixel co-ordinates. In a structured light system, the projected image is perfectly known because there is no error on the pixel co-ordinates measurement. Error is only added on the pixel co-ordinates of the camera. The reconstruction was performed on 40 points. Table 1 presents the result for 10 of these 40 points with a uniform noise on pixel within ±1. It has to be noted that using five known 3-D points results degrade quickly. This method appears to be very sensitive to the location of the five points used as landmarks.

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

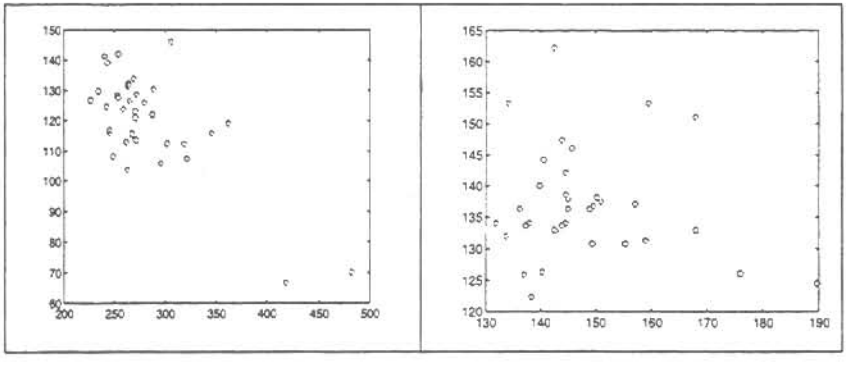

Figure 3: Validation of the projective reconstruction, left and right image planes. Circles represent real image points and crosses represent projectively reconstructed ones.

5.2. Real data

We have reconstructed 70 points of a scene, which represents a cube (see figure 4a). 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 4d. 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 show that uncalibrated reconstruction from grid coding performed well (in a qualitative way) and that its accuracy is good enough to be used in mobile robotics. On figure 5, we present results of Euclidean reconstruction performed from an image grabbed under realistic condition. Only the highlighted lines have been reconstructed.

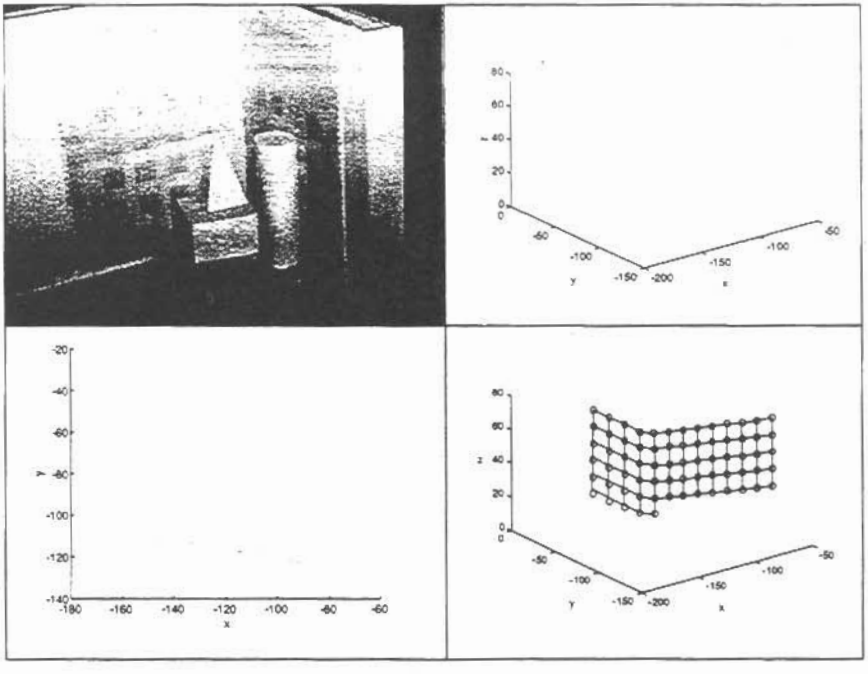

Figure 4: Euclidean Reconstruction by geometrical constraints: (a) Grey-level image (b and c) Two views of an Euclidean reconstruction (d) Comparison with hard-calibrated reconstruction in circles

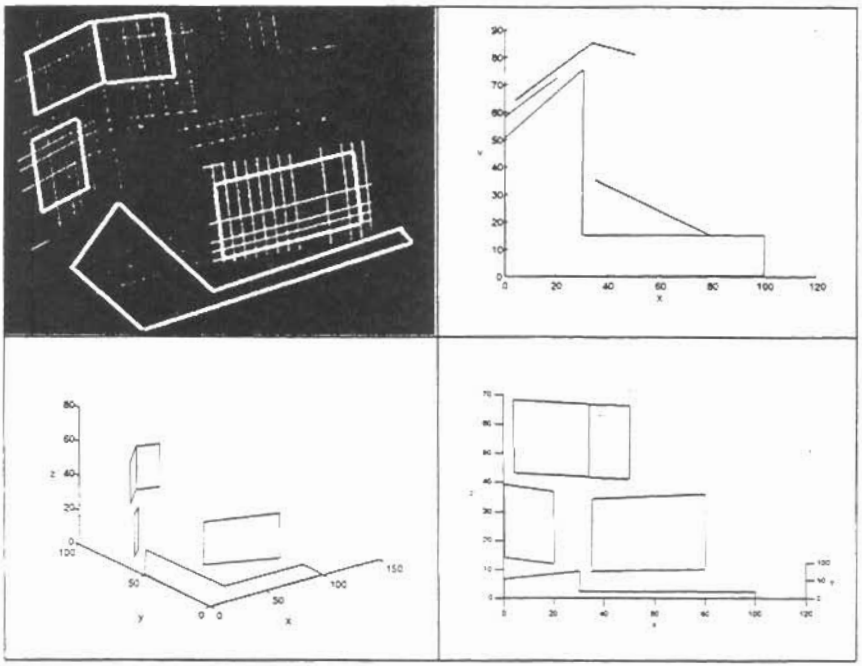

Figure 5: Euclidean Reconstruction by geometrical constraints: (a) Structured light image (b, c and d) Three views of an Euclidean reconstruction

6. CONCLUSIONS

This article presents a method to perform Euclidean reconstruction from an uncalibrated structured light sensor. Through pixel correspondences and without knowing neither extrinsic nor intrinsic parameters of the sensor, a projective reconstruction is first computed by choosing five arbitrary points of the scene as a reference frame. Such a reconstruction is only possible up to a projective transformation, which depends on the choice of the five points. Since Euclidean geometry is a particular case of projective geometry, a collineation exists which brings projective reconstruction to Euclidean reconstruction. This collineation can be assessed by translating geometrical information about the scene into constraints on the entries of the collineation matrix. Besides, we show that projecting a known pattern onto the scene permits to retrieve intrinsic geometrical knowledge about this scene as parallelism or orthogonality. It is important to note that the method based on fixing Euclidean constraints and the adaptation to grid coding are independent from projective reconstruction. As no constraint is required on projection matrices, this approach allows us to reconstruct changing the focus, the aperture and the zoom

of both the camera and the projector. Structured lighting permits to ensure there is known scene structure which can be used to upgrade the reconstruction to Euclidean and provides numerous constraints which are useful for the convergence of non-linear optimisation methods as Levenberg-Marcquardt algorithm.

Experimental results validate the method, although further works have to be undertaken. In order to use this reconstruction method in mobile robotics, the automation of the whole of the process is necessary. Constraints have to be automatically generated, from their extraction to their mathematical formulation. So, a particular care has to be taken in image segmentation, *e.g.* straight lines, parallels, crossing lines and parallelograms must be accurately extracted from the image.

References

[1] J. Batlle, E. Mouaddib, J. Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem. A survey", *International Journal of Pattern Recognition*, 31(7), pp. 963-982, 1998

[2] B. Boufama, R. Mohr, F. Veillon, "Euclidean constraints for uncalibrated reconstruction", *Proceedings of the 4th International Conference on Computer Vision*, Berlin (Germany), pp. 466-470, may 1993.

[3] O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig ?", *Proceedings of the 2nd European Conference on Computer Vision*, Santa Maria Ligure, Italia, pp. 563-578, 1992.

[4] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*, MIT Press, 1993.

[5] E. L. Hall, J. B. K. Tio, C. A. McPherson, C. S. Draper, F. A. Sadjadi, "Measuring Curved Surfaces for Robot Vision", *International Journal on Computer Vision*, Vol. December, pp. 42-54, 1982.

[6] R. Hartley, R. Gupta, T. Chang, "Stereo from uncalibrated cameras", *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Urbana-Champaign, Illinois (USA), pp. 761-764, 1992.

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

[8] R. Mohr, B. Boufama, P. Brand, "Accurate projective reconstruction", *Proceedings 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", *International Journal of Pattern Recognition Letters*, Vol. 19, pp. 1055-1065, 1998.

[10] G. Sparr. "Projective invariants for affine shapes of point configurations", *Proceeding og the DARPA-ESPRIT Workshop on Applications of Invariants in Computer Vision*. Reykjavik (Island), pp. 151-170, march 1991.

[11] D. Weinshall, "Model-based invariants for 3-D vision", *International Journal of Computer Vision*, 10(1), pp. 27-42, 1993.

[12] Z. Zhang, G. Xu, "A general expression of the fundamental matrix for both perspective and affine cameras", *Proceedings of the 15th International Joint Conference on Artificial Intelligence*, Nagoya (Japan), august 1997.