PERGAMON

Pattern Recognition 36 (2003) 1631–1644

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Uncalibrated reconstruction: an adaptation to structured light vision

David Fofi[a], Joaquim Salvi[b], El Mustapha Mouaddib[a,*]

[a] *CREA, Université de Picardie Jules Verne, 7 rue du Moulin Neuf, 80000 Amiens, France*
[b] *Computer Vision and Robotics Group, IIiA, Universitat de Girona, Avda. Lluis Santalo, s/n, 17071 Girona, Spain*

Received 15 May 2002; accepted 23 September 2002

Abstract

Euclidean reconstruction from two uncalibrated stereoscopic views is achievable from the knowledge of geometrical constraints about the environment. Unfortunately, these constraints may be quite difficult to obtain. In this paper, we propose an approach based on structured lighting, which has the advantage of providing geometrical constraints independent of the scene geometry. Moreover, the use of structured light provides a unique solution to the tricky correspondence problem present in stereovision. The projection matrices are first computed by using a canonical representation, and a projective reconstruction is performed. Then, several constraints are generated from the image analysis and the projective reconstruction is upgraded into an Euclidean one—as we will demonstrate, it is assumed that the sensor behaviour is affine without loss of generality so that the constraints generation is simplified. The method provides our sensor with adaptive capabilities and permits to be used in the measurement of moving scenes such as dynamic visual inspection or mobile robot navigation. Experimental results obtained from both simulated and real data are presented.

Keywords: Uncalibrated system; Projective reconstruction; Euclidean constraints; Structured light; Computer vision

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 challenge is to infer 3D information of a scene by starting from at least two images of it [1]. From two images, the reconstruction is thus possible but we need to calibrate the cameras, 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). The correspondence among the 3D object points and their projections has to be established; the obtained matrix makes possible to relate each point to its line of sight [2]. This technique is named *hard-calibration* and is carried out off-line by using a calibrating pattern whose 3D 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 the measurement of moving scenes such as dynamic visual inspection or mobile robot navigation. Then, the visual adaptation permits to use 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.

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

* Corresponding author. Tel./fax: +33-3-22-82-76-68.
E-mail address: mouaddib@u-picardie.fr (E.M. Mouaddib).

0031-3203/03/$30.00
PII: S0031-3203(02)00288-1

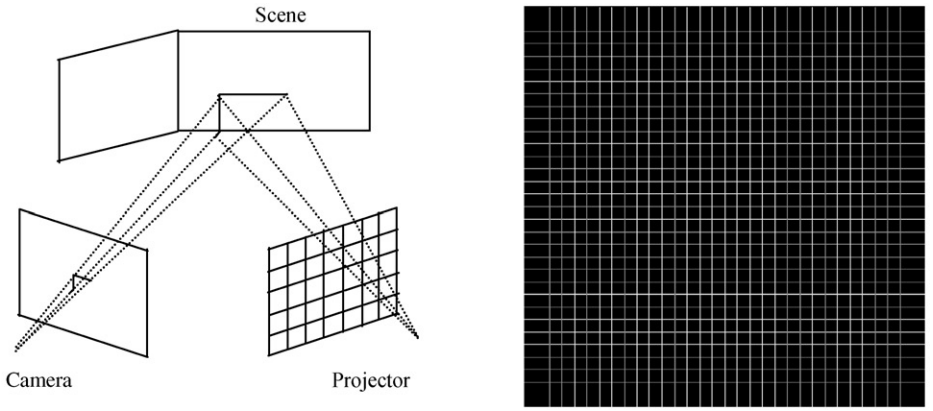

Fig. 1. The structured light system: (a) the geometrical principle; (b) the colour-encoded pattern.

this problem, coded structured light techniques have been developed [3]. 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 Fig. 1a. Since a projector can be seen as a camera acting in reverse, it can be modeled in the same way a camera is.

Our pattern is composed by a set of vertical and horizontal slits, uniquely colour-encoded in a single pattern projection (Fig. 1b). The reader is pointed to Salvi et al. [4] to get deeper into the pattern design. The coloured codification permits to solve the correspondence problem finding out for each imaged point, its corresponding point in the projecting plane.

The main goal of this paper is, firstly, to provide adaptive capabilities to our structured light vision sensor and, secondly, to adapt the techniques of uncalibrated reconstruction to structured light. Our contribution 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 of the scene and, more generally, to show how to self-calibrate a structured light sensor. It is assumed throughout the paper that the camera and the projector behaviour can be approximated by an affine model. However, no assumption on the scene geometry is imposed, although a planar piece-wise environment ensures a major number of constraints.

The next points summarize our approach:

- Extraction of the image points by a specific image processing (see the section experimental results).
- Projective reconstruction from one view and one pattern projection using the canonical representation.
- Automatic generation of constraints in order to reach an Euclidean reconstruction of the scene.

This article is organized as follows. Section two presents the related work about reconstruction methods that could be adapted to a structured light system. Then, section three details the Euclidean reconstruction through structured lighting, which contains the major contribution of this paper. Furthermore, section four deals with some experimental results considering both simulated and real scenes. The article ends with conclusions.

2. Reconstruction and structured light system

This section goes deeper into what has been proposed on structured light, taking advantages of the specificities of projection and pattern structure and considering that the principle of calibration is known. For instance, Salvi et al. [4] proposed to model the projector like a camera acting in reverse. First, they calibrate the camera by using a calibrating plane and then an image of the projected pattern on the calibrating plane is grabbed and used to get the 3D points to calibrate the projector. Proesmans et al. [5] proved that the reconstruction could be performed whether the angle between directions of projection and capture is known, assuming an orthographic model. This particular way to calibrate consists in observing a blank calibration pattern whose angle, made up by the two planes that compose it, is precisely known. Sotoca et al. [6] proposed a calibration method adapted to large surface. Beforehand, the pattern is projected onto a *base plane* and onto a *reference plane* and an image is grabbed for each of these planes. By positioning the object to analyze between these planes, the authors show that it is possible to obtain a depth image through some simple calculations based on the similarity of triangles. Finally, let us conclude by the method developed by Huynh et al. [7] which has been proposed for light plane projections but it can be generalized to pattern projections. Four sets of three coplanar points, whose coordinates are precisely known, are positioned on the two planes of a calibration pattern, depicting four lines on it. While the light plane intersects these lines, a fourth point is obtained on each of them: the cross-ratio of these points is equal to the cross-ratio of the images of these points. It

provides a 3D measurement of the lighted traces fairly precise (up to the stability of cross-ratios). A classical calibration process is performed from these measurements.

Uncalibrated vision has generated an increasing number of publications since the end of the 1980s. Aware of the drawbacks of hard-calibration for some applications in which the sensor has to adapt its behaviour to the variations of the environment and to the strategy of observation, many authors looked into the problem that consists of inferring 3D structure of the scene from the pixel coordinates only. It is proved that an Euclidean reconstruction cannot be obtained without calibrating but, at best, a projective one [2]; this reconstruction has to be constrained by means of additional information or assumptions (about the intrinsic parameters, the movement of the sensor, the scene geometry) to achieve the self-calibration of the sensor, i.e. the Euclidean reconstruction of the scene.

Koenderink and Van Doorn [8] can be regarded as pioneers. In 1989, they proposed a method which allows to recover the affine scene structure from at least two images of it, by using a shape invariant computed from a reference plane. Later, Faugeras [2] and, independently, Hartley et al. [9] proved that from a weakly calibrated sensor (i.e. which epipolar geometry is known) a projective reconstruction is possible. Mohr et al. [10] reached similar results through a global estimation of the unknowns by minimizing the residual errors between image points and their back-projections. However, these methods do not allow to compute the Euclidean structure of the environment. With this aim, Faugeras et al. [11] proposed a method that takes advantage of the invariance under similarities of the absolute conic (in other words, the image of the conic only depends on intrinsic parameters of the camera). By such performing, the authors have rediscovered the Kruppa equations. Besides, Hartley [12] demonstrated that these equations could be explicitly obtained through the singular value decomposition of the fundamental matrix $\mathbf{F}$. In order to solve them, three fundamental matrices have to be computed, which are given from a single displacement of the stereo head.

A second class of methods assumes that intrinsic parameters remain constant during the measuring process. A projective reconstruction is first performed. Then, the constancy assumption leads to an equation solvable from three views of the scene; the reader can refer to the work of Hartley [13] or Heyden and Aström [14]. At last, if the constancy of intrinsic parameters cannot be assumed, it is possible to upgrade a projection reconstruction into an Euclidean one by generating Euclidean constraints grabbed from the scene geometry. Boufama et al. [15] pioneered this method; Zhang et al. [16] later on proposed a similar method.

Some considerations have to be taken in mind when structured light is used. Any movement of the sensor, and particularly of the projector, produces a sliding of the projected points on the observed surfaces. That is to say, the points illuminated before the movement are different than the ones illuminated after the movement. As a consequence, stereovision algorithms using more than two views cannot be adapted to structured light vision. Besides, due to the heterogeneity of the sensor, composed by a camera and a projector, the constancy of intrinsic parameters cannot be assumed either. Hence, methods based on Kruppa's equations and methods based on constant intrinsic parameters are unsuited to structured light vision.

There is only one choice left: performing a projective reconstruction first and rectifying it by using Euclidean constraints grabbed from the scene geometry. It is shown in the next section what kind of projective reconstruction method may be used and how to generate constraints by using the geometry of light patterns.

3. Uncalibrated reconstruction and structured light system

This section details a method that permits to locate a point in the three-dimensional space from a pair of uncalibrated perspective views (which is equivalent to one view from a camera and a known projected pattern). First, the method performs a reconstruction in a projective frame. Then, the reconstruction of the scene is transformed to an Euclidean frame by using some a priori knowledge between the view of the scene and the projected pattern, less restrictive than point co-ordinates, such as parallelism, orthogonality, angles, length ratio, and so on.

The following section presents the theoretical basis and methods about the uncalibrated reconstruction adapted to structured light. Then, Section 3.5 details the proposed algorithm that describes the whole process.

3.1. Projective reconstruction

It is known, since the work of Luong and Vieville about the canonical representation of the geometry of multiple views [17], that it is possible to estimate the camera projection matrices from the knowledge of epipolar geometry. Then, considering two 2D images (an image of the scene and the projected pattern), we have:

$$\mathbf{P}_{proj} = [\mathbf{M}\ \ \mathbf{e}], \quad \mathbf{P}'_{proj} = [\mathbf{I}\ \ \mathbf{0}]$$

$$\text{with } \mathbf{M} = -\frac{1}{\|\mathbf{e}\|^2}[\mathbf{e}]_\times \mathbf{F}, \tag{1}$$

where $\mathbf{F}$ denotes the fundamental matrix and $\mathbf{e}$ the epipole of the first image, $\mathbf{P}$ is the camera matrix and $\mathbf{P}'$ the projector matrix. Subscript *proj* denotes matrices, vectors or scalars expressed in a projective frame, in contrast with *eucl* which will denote matrices, vectors or scalars expressed in an Euclidean frame.

Whereas $\mathbf{F}^{\mathrm{T}}\mathbf{e} = 0$, so that the coordinates of the epipole are given by the eigenvector of the matrix $\mathbf{F}\mathbf{F}^{\mathrm{T}}$ associated with the smallest eigenvalue. Numerically, better results are obtained by normalizing the epipole in the way that $\|\mathbf{e}\| = 1$.

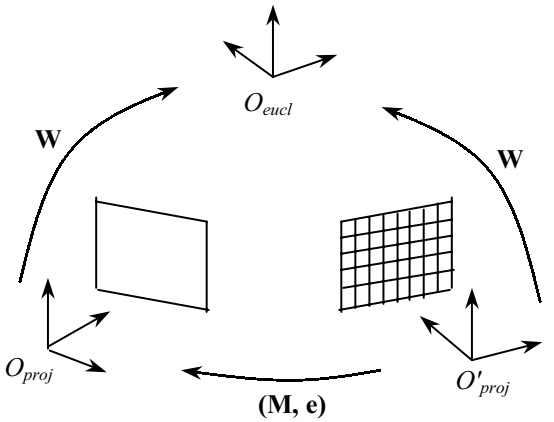

Fig. 2. Reference frames and projection matrices for the projective reconstruction.

This formulation satisfies the epipolar constraint and fixes the projective basis on the camera frame; where the projection matrices are computed so that the 3D observed points could be reconstructed with respect to *that* frame. We will perform it in two steps: first, by a linear method, fast but not accurate enough; then, by a non-linear method initialized with the results of the linear method. We describe the two steps in the following notations and reference frames are depicted in Fig. 2.

3.1.1. Linear method

Given a pair of points in correspondence, $\mathbf{m} = [u \;\; v \;\; 1]^{\mathrm{T}}$ and $\mathbf{m}' = [u' \;\; v' \;\; 1]^{\mathrm{T}}$, and their corresponding 3D point in space $\mathbf{M}_{proj} = [x \;\; y \;\; z \;\; t]^{\mathrm{T}}$ expressed in the projective frame, it is obtained:

$$\lambda[u \quad v \quad 1]^{\mathrm{T}} = \mathbf{P}_{proj}[x \quad y \quad z \quad t]^{\mathrm{T}} \tag{2}$$

$$\lambda'[u' \quad v' \quad 1]^{\mathrm{T}} = \mathbf{P}'_{proj}[x' \quad y' \quad z' \quad t']^{\mathrm{T}} \tag{3}$$

where λ and λ' are two non-zero scale factors. Eliminating the scale factors and re-arranging equations (2) and (3) yields to

$$\mathbf{Q}\mathbf{M}_{proj} = 0. \tag{4}$$

$\mathbf{Q}$ is a 4×4 matrix given by

$$\mathbf{Q} = [\mathbf{p}_1 - u\mathbf{p}_3 \quad \mathbf{p}_2 - v\mathbf{p}_3 \quad \mathbf{p}'_1 - u'\mathbf{p}'_3 \quad \mathbf{p}'_2 - v'\mathbf{p}'_3], \tag{5}$$

where $\mathbf{p}_i$ and $\mathbf{p}'_i$ are the vectors corresponding to the ith row of $\mathbf{P}$ and $\mathbf{P}'$, respectively. As $\mathbf{M}_{proj}$ is defined up to a scale factor, we can impose $\|\mathbf{M}\| = 1$. The solution is given by the eigenvector of the matrix $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$ associated to the smallest eigenvalue.

3.1.2. Non-linear method

It seems difficult, in the previous approach, to give a good physical interpretation to the criterion that is minimized. Besides, the accuracy of the results can be significantly improved. A way to alleviate these drawbacks is to use a non-linear iterative method of minimization. The error to minimize is the difference between the observation and the back-projection of the reconstructed points or *residual* error; in other words:

$$\left(u - \frac{\mathbf{p}_1^{\mathrm{T}}\mathbf{M}_{proj}}{\mathbf{p}_3^{\mathrm{T}}\mathbf{M}_{proj}}\right)^2 + \left(v - \frac{\mathbf{p}_2^{\mathrm{T}}\mathbf{M}_{proj}}{\mathbf{p}_3^{\mathrm{T}}\mathbf{M}_{proj}}\right)^2 + \left(u' - \frac{\mathbf{p}_1'^{\mathrm{T}}\mathbf{M}_{proj}}{\mathbf{p}_3'^{\mathrm{T}}\mathbf{M}_{proj}}\right)^2 + \left(v' - \frac{\mathbf{p}_2'^{\mathrm{T}}\mathbf{M}_{proj}}{\mathbf{p}_3'^{\mathrm{T}}\mathbf{M}_{proj}}\right)^2. \tag{6}$$

In practice, a traditional algorithm of minimization like Levenberg–Marquardt is used [18]. The results provided by the direct method are used to initialize the algorithm.

3.2. Towards an Euclidean reconstruction

It is known that the Euclidean geometry is a particular case of the projective geometry. In other words, a collineation exists which brings the solution to an Euclidean one. Finding this collineation, it is thus possible to recover the Euclidean structure of the scene.

Let us consider $\mathbf{M}_{proj}$, a point with projective co-ordinates, and $\mathbf{M}_{eucl}$ the same point with Euclidean co-ordinates.

$$\mathbf{M}_{proj} = \begin{pmatrix} x_{proj'} \\ y_{proj} \\ z_{proj} \\ t_{proj} \end{pmatrix}, \quad \mathbf{M}_{eucl} = \begin{pmatrix} x_{eucl} \\ y_{eucl} \\ z_{eucl} \\ 1 \end{pmatrix}. \tag{7}$$

The problem is to determine $\mathbf{W}$ such as

$$\mathbf{M}_{eucl} = \mathbf{W} \cdot \mathbf{M}_{proj}. \tag{8}$$

With the aim of computing the collineation, geometrical knowledge about the scene is translated into constraints on the entries of $\mathbf{W}$.

We have to fix affine or Euclidean constraints, which reports geometrical properties of the scene extracted from the images. $\mathbf{W}$ has 15 degrees of freedom; therefore 15 independent and coherent constraints have to be found. Hereafter, a non-exhaustive list of constraints with their mathematical formulations is given and it is shown how the use of structured light leads to generate such constraints.

3.3. Euclidean constraints from grid coding

As described in Section 1, grid coding is the way a grid of light is projected onto the scene to be analyzed. It is the sub-class of structured light techniques that used a grid pattern. It is shown in this section how geometrical knowledge about the scene can be obtained analyzing the grid-coded images. This knowledge can be used as constraints to bring a projective reconstruction to an Euclidean one.

3.3.1. 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

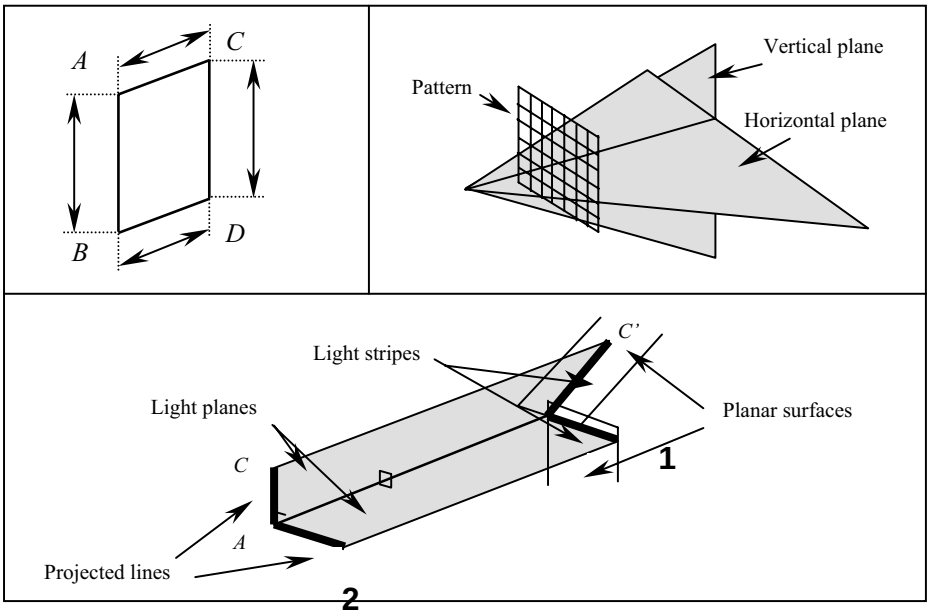

Fig. 3. Some of the geometrical constraints used in Euclidean reconstruction: (a) Parallelogram constraint; (b) Horizontal and vertical plane; (c) Orthogonality.

quadrilateral. Quadrilaterals can theoretically either be captured as squares, rectangles, rhombuses, parallelograms or trapezoids, depending on the position and orientation of the camera and projector, and the planar surface observed. Although quadrilateral detection within the image is not equivalent to plane detection in 3D space (there are configurations of curved surfaces that can yield quadrilaterals in the image), it is quite probable that a quadrilateral within the image corresponds to a plane in space: the equivalence is assumed throughout the paper and, moreover, this can be verified, thanks to the test of coplanarity we propose in Section 4.1.

3.3.2. *Parallelogram constraints*

Assuming that the projector have an approximately affine behaviour, we obtain that if a square is projected onto a planar surface, the more generic quadrilateral formed on the surface 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 3D plane. Geometrical knowledge about the scene can thus be deduced.

Relative positioning of the four points A, B, C and D of the parallelogram (see Fig. 3a) in space is such as

$$\overline{AB} = \overline{CD}, \quad \overline{AC} = \overline{BD}, \tag{9}$$

$$(AB)//(CD), \quad (AC)//(BD). \tag{10}$$

It leads to a *redundant* set of constraints on W. Besides, knowing Eq. (11), parallelism constraints can be simplified as shown in Eqs. (11) and (12):

$$(AB)//(CD) \Leftrightarrow \frac{\overrightarrow{AB}}{\|\overrightarrow{AB}\|} = \frac{\overrightarrow{CD}}{\|\overrightarrow{CD}\|}, \tag{11}$$

$$\begin{aligned}&(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2\\ &\quad = (x_D - x_C)^2 + (y_D - y_C)^2 + (z_D - z_C)^2,\\ &(x_C - x_A)^2 + (y_C - y_A)^2 + (z_C - z_A)^2\\ &\quad = (x_D - x_B)^2 + (y_D - y_B)^2 + (z_D - z_B)^2,\end{aligned} \tag{12}$$

$$\begin{aligned}&(x_B - x_A) = (x_D - x_C), (y_B - y_A) = (y_D - y_C), (z_B - z_A)\\ &\qquad = (z_D - z_C),\\ &(x_C - x_A) = (x_D - x_B), (y_C - y_A) = (y_D - y_B), (z_C - z_A)\\ &\qquad = (z_D - z_B).\end{aligned} \tag{13}$$

Since projective geometry keeps unchanged the alignment and the coplanarity, Eqs. (12) and (13) determine the same configuration of points (redundant constraints). Note that a parallelogram completely determines a 3D plane. Therefore, for each plane composing the scene, a unique set of parallelogram constraints is sufficient.

Now, let us consider the two configurations of points shown in Fig. 4. Whether the points o_i, p_i, q_i, r_i and s_i ($i = 1$ or 2) are projected onto a plane, the cross-ratio within the pattern is equal to the c ross-ratio of the five points formed onto this plane; moreover, the cross-ratio of the homologue points within the image is equal to both. The change from projected points to imaged points is obtained by two successive homographies. It can be deduced that if

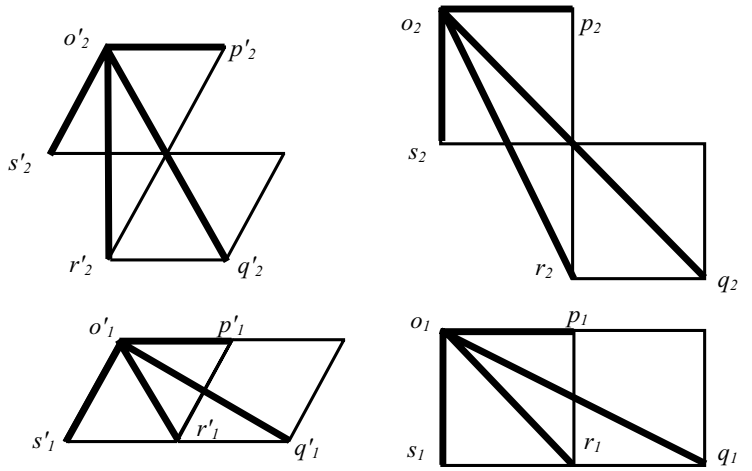

Fig. 4. Test of coplanarity. Points configuration: image plane (left) and projector plane (right).

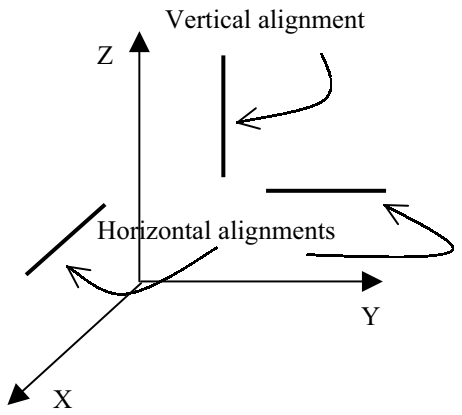

Fig. 5. Euclidean constraint: alignment.

Eq. (14) is verified, the corresponding object points O, P, Q, R and S are coplanar.

$$\{o_i; p_i, q_i, r_i, s_i\} = \{o_i'; p_i', q_i', r_i', s_i'\} \quad \text{with } i = 1 \text{ or } 2. \tag{14}$$

3.3.3. 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 the following linear constraint (Fig. 5):

$$w_{31}x_{P'} + w_{32}y_{P'} + w_{33}z_{P'} + w_{34}t_{P'} = 0. \tag{15}$$

Replacing w_{3i} by w_{2i} or w_{1i}, the homologue constraints for $y_P = 0$ or $x_P = 0$, respectively, are expressed. Noting that each projected horizontal line of the pattern generates a light plane in space, which can be considered as a 3D horizontal plane in the projector co-ordinate system (see Fig. 3b). And each projected vertical line of the pattern generates a light plane in space, which can be considered as a vertical 3D 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 the projecting planes of light with the scene surfaces, therefore points belong to horizontal or vertical planes.

Furthermore, an arbitrary distance can be set between two successive horizontal or vertical planes. If the distance between two points A and B is assumed to be d. Then, $(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 = d^2$ and, as a consequence, the following non-linear constraint is obtained:

$$\left(\frac{w_{11}x_{A'} + w_{12}y_{A'} + w_{13}z_{A'} + w_{14}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} - \frac{w_{11}x_{B'} + w_{12}y_{B'} + w_{13}z_{B'} + w_{14}t_{B'}}{w_{41}x_{B'} + w_{42}y_{B'} + w_{43}z_{B'} + w_{44}t_{B'}}\right)^2 + \left(\frac{w_{21}x_{A'} + w_{22}y_{A'} + w_{23}z_{A'} + w_{24}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} - \frac{w_{21}x_{B'} + w_{22}y_{B'} + w_{23}z_{B'} + w_{24}t_{B'}}{w_{41}x_{B'} + w_{42}y_{B'} + w_{43}z_{B'} + w_{44}t_{B'}}\right)^2 + \left(\frac{w_{31}x_{A'} + w_{32}y_{A'} + w_{33}z_{A'} + w_{34}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} - \frac{w_{31}x_{B'} + w_{32}y_{B'} + w_{33}z_{B'} + w_{34}t_{B'}}{w_{41}x_{B'} + w_{42}y_{B'} + w_{43}z_{B'} + w_{44}t_{B'}}\right)^2 = d^2. \tag{16}$$

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

Without particular knowledge, a plane can be arbitrarily chosen as a horizontal or vertical plane; in this case, the

reconstruction will be performed up to a rotation and a translation along the z-axis.

3.3.4. Fixing the origin

If the Euclidean co-ordinates of a point p in space are known, it is obtained that

$$x_p = \frac{w_{11}x_{p'} + w_{12}y_{p'} + w_{13}z_{p'} + w_{14}t_{p'}}{w_{41}x_{p'} + w_{42}y_{p'} + w_{43}z_{p'} + w_{44}t_{p'}},$$

$$y_p = \frac{w_{21}x_{p'} + w_{22}y_{p'} + w_{23}z_{p'} + w_{24}t_{p'}}{w_{41}x_{p'} + w_{42}y_{p'} + w_{43}z_{p'} + w_{44}t_{p'}}$$

$$z_p = \frac{w_{31}x_{p'} + w_{32}y_{p'} + w_{33}z_{p'} + w_{34}t_{p'}}{w_{41}x_{p'} + w_{42}y_{p'} + w_{43}z_{p'} + w_{44}t_{p'}}. \quad (17)$$

Then, these equations give three linear constraints. As we said before, this knowledge is barely available; nevertheless, we can fix the origin of the Euclidean co-ordinates frame by equaling these equations to zero. The cross-point (which appears in the image as the intersection of two light stripes) of the planes $y = 0$ and $x = 0$ is considered to be the origin of the Euclidean co-ordinate frame.

3.3.5. Orthogonality constraint

Orthogonality is an important feature in Euclidean reconstruction. The detection of orthogonal planes permits to define, at least partially, a 3D 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 (Fig. 3c). When the projecting light planes intersect planar surfaces, they produce light stripes on them which 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, considering the world co-ordinate system is fixed at the projector, it is obtained

$$x_{A'} = x_{B'}, \quad y_{A'} = y_{C'}, \quad (18)$$

$$\begin{aligned}\overrightarrow{A'B'} \cdot \overrightarrow{A'C'} &= (x_{A'} - x_{B'})(x_{A'} - x_{C'}) \\ &\quad + (y_{A'} - y_{B'})(y_{A'} - y_{C'}) \\ &\quad + (z_{A'} - z_{B'})(z_{A'} - z_{C'}) \\ &= (z_{A'} - z_{B'})(z_{A'} - z_{C'}). \quad (19)\end{aligned}$$

So

$$(A'B') \perp (A'C') \iff z_{A'} = z_{B'} \text{ or } z_{A'} = z_{C'}. \quad (20)$$

If the conditions imposed by (21) are satisfied, we obtain an *orthogonality constraint*, otherwise we obtain a *reduced orthogonality constraint*:

$$(x_{A'} - x_{B'})(x_{A'} - x_{C'}) + (y_{A'} - y_{B'})(y_{A'} - y_{C'}) = 0. \quad (21)$$

3.4. Resolution of the system

The projective reconstruction is first performed by solving the set of equations (1) previously described. In a least-squares optimization, it leads to minimize the following error:

$$(\hat{A}, \hat{P}) = \arg\min_{A,P} (p - AP)^{\mathrm{T}} C^{-1} (p - AP), \quad (22)$$

where $\hat{A}$ and $\hat{P}$ are the estimated values of A and P, respectively; and C is the covariance matrix. As the location of points within the images is the major factor of noise, all the other factors are neglected. Then, C is a diagonal matrix and its elements are all equal to the variance since imprecise location induces decorrelated noise.

The Levenberg–Marquardt algorithm [18] is used to solve this set of non-linear equations. As 3D points and projective matrices can only be known up to a scale factor, a scale constraint must be added for each point and each matrix in order to lead to a unique solution. Eq. (23) is the constraint for points and Eq. (24) for matrices:

$$x_i^2 + y_i^2 + z_i^2 + t_i^2 - 1 = 0, \quad (23)$$

$$m_{34}^{(j)} = 1, \quad j = 1, \ldots, s. \quad (24)$$

Once the projective reconstruction is performed, the matrix W has to be estimated to obtain the Euclidean reconstruction. The Levenberg–Marquardt algorithm is also used. Equations which should be minimized are the ones that provide Euclidean constraints (Eqs. (14)–(18) and (22)). The scale constraint that has to be added is given by Eq. (25).

$$\sum_{i,j} (w_{ij})^2 = 1. \quad (25)$$

3.5. Algorithm

Let us now summarize the steps that are necessary to perform an Euclidean reconstruction without any a priori knowledge about the observed scene, but a single image.

1. Image processing
Input: camera image and projected pattern
Output: the two sets of matched points (U_{ij}, V_{ij}, the co-ordinates of the ith point in the jth image)

- Segmentation and decoding.
- Solve the correspondence problem decoding the pattern.

2. Projective reconstruction (Section 3.1):
Input: matching points
Output: 3D reconstructed points in a projective frame

- Estimation of the fundamental matrix.
- Estimation of the projection matrices.
- 3D reconstruction by linear method.

- 3D reconstruction by iterative method (using the previous method as initialization).

3. Euclidean reconstruction:
Input: 3D projective points and projection matrices, segmented and decoded lines and cross-points.
Output: 3D reconstructed points in a Euclidean frame.

- Fix an arbitrary point as the origin of the world co-ordinate system (Section 3.3.4)
- Fix the horizontal (in the pattern) line to which it belongs as the $y = 0$ plane in space (Section 3.3.3).
- Fix the vertical (in the pattern) line to which it belongs as the $x = 0$ plane in space (Section 3.3.3).
- Extract parallelograms within the image and generate parallelogram constraints, if the test of coplanarity is ok (Section 3.3.2).
- Extract crossing lines within the image and generate (reduced) orthogonality constraints (Section 3.3.5).
- Fix an arbitrary distance d between two points in space (Section 3.3.5).
- Compute the collineation in order to upgrade the projective reconstruction into Euclidean one (Section 3.2).

4. Experimental results

First of all, the stability of cross-ratio is discussed in order to evaluate the efficiency of our test of coplanarity and some results are presented. Then, experiments on reconstruction have been performed with simulated data. A set of 3D points is fixed with respect to a world co-ordinate system and these points are observed by two virtual cameras. Five of these points are chosen as the reference basis; all the other ones are reconstructed using the method previously described, i.e. by using only their pixel coordinates and the matching in both images. Euclidean reconstruction by adding geometrical constraints is obtained and results validated. Furthermore, we have performed the reconstruction method with real images using our structured light sensor. In the following, we detailed the implementation of the algorithm and we summarize the experimental results obtained. Quantitative results are given for the experiments performed with simulated data; since we do not have reliable measurements of the real scenes, only qualitative results are given for the experiments with real data.

All the experiments have been performed by using Matlab, so time consuming is not very significant. The test of coplanarity and the linear method for projective reconstruction, based on matrix algebra, are achieved in less than one second. In contrast, the Euclidean reconstruction (i.e. the Levenberg–Marquardt algorithm) is performed in a few iterations (from 5 to 20, depending on the Euclidean constraints), that is, in a few minutes. Of course, by programming the algorithm in C/C++ code, it is possible to considerably improve time consuming (a set of non-linear equations can be solved in less than one second in C/C++).

4.1. Test of coplanarity

We have tested the stability of the cross-ratio for the configurations of points required by the test of coplanarity. We took five points separated by the distance d (on Fig. 4, d is the distance o_1p_1, o_2p_2, etc.) A gaussian noise, varying from 0 to $0.5 \times d$, is added on the points co-ordinates. The results are depicted by Fig. 6. The left part shows the stability of cross-ratio with a noise of $\pm 5\%$ for 100 computed values (theoretically, cross-ratio is 2 in this example). The right part shows the evolution of the error against the noise level (which depends on d).

To be able to compare the theoretical cross-ratios with the cross-ratios computed from the images (i.e. to compute the error) we used a *projective distance* based on the method of the random cross-ratios, detailed in Ref. [19]. 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 coplanar points. Obviously, as it can be deduced from the results of Fig. 6: the larger the distance d is, the more robust the measure of cross-ratio will be. The left part of Fig. 6 shows that, with a moderate noise ($\pm 5\%$), the measured cross-ratio is very near to the theoretical one. Hence, the stability of cross-ratio is good enough for applications of uncalibrated reconstruction.

We have tried out the test of coplanarity by performing three experiments. In the first one, a planar configuration of points is detected (theoretical cross−ratio $= 2$, measured cross−ratio $= 1.96$, projective error $= 2.2 \times 10^{-3}$). In the second one, the pattern is projected onto an irregular surface and the test classifies these points as non-coplanar (theoretical cross−ratio $= 2$, measured cross−ratio $= 2.186$, projective error $= 5.9 \times 10^{-3}$). Finally, in the last experiment, the points are projected onto a cube corner (clearly not coplanar) and the points are well-classified (theoretical cross−ratio$=2$, measured cross−ratio$=2.2055$, projective error $= 9.9 \times 10^{-3}$). As this test is only based on cross-ratio computing, its time computing is near-instantaneous.

4.2. Simulated data

4.2.1. Five known points

It is assumed here that five points of the scene are taken as landmarks whose Euclidean co-ordinates are known. Let us assume that the camera is set at the origin of the world co-ordinate system. Only four independent parameters have to be estimated in order to obtain the projection matrix of the camera. The co-ordinates of the principal point are initialized with the co-ordinates of the geometrical image centre. The 3D point co-ordinates are initialized as the co-ordinates of the barycentre of the points to be reconstructed. Obviously, with simulated data and no noise, the discrepancy between

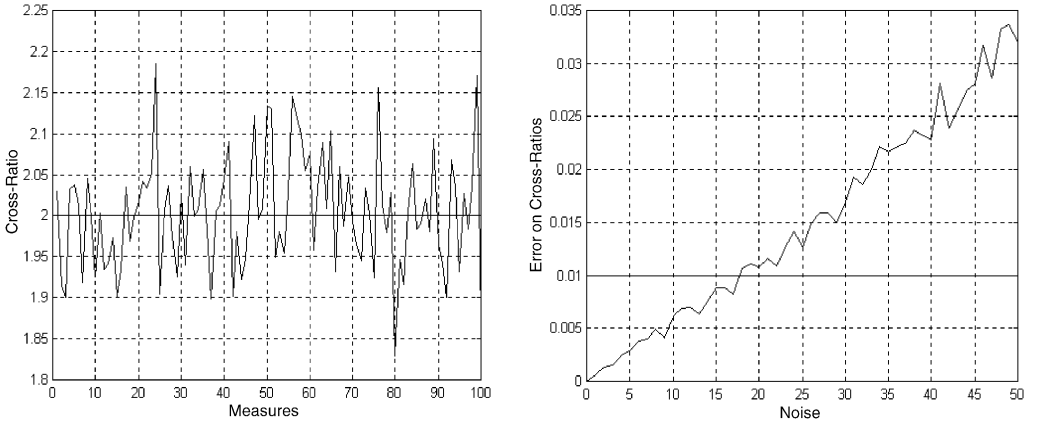

Fig. 6. Cross-ratio stability. 100 computations with a 5% noise level (left) and measuring error with a noise varying from 0% to 50% (right).

the real 3D points and the reconstructed 3D points is nearly zero. The error is due to round off in digital computation. So, in order to analyze the robustness of the method, noise is added on pixel co-ordinates. In a structured light system, the projected image is perfectly known so that error is not present when the point co-ordinates are measured. Then, noise 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 of $\pm 1\%$.

It has to be noticed that using five known 3D points results degrade quickly. This method appears to be very sensitive to the location of the five points used as landmarks: better results are obtained if no noise is added on landmarks co-ordinates even if noise is added on the co-ordinates of the rest of points.

4.2.2. Euclidean constraints

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. Projective reconstruction provides projection matrices and 3D co-ordinates with respect to a projective frame. In order to validate this reconstruction, the 3D co-ordinates are back-projected onto the image planes through the projection matrices and the residual error is evaluated (see Fig. 7, where projection parameters are given by $\mathbf{P}_{proj}$ and $\mathbf{P}'_{proj}$). 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 3D points must correspond more or less to the configuration of the chosen basis i.e. the Euclidean reference points must be in adequacy with the projective co-ordinates given to them.

We have used different Euclidean constraints as fixing the origin, parallelism, distance, etc. Re-scaling and re-positioning the computed reconstruction, it is possible to validate Euclidean reconstruction. In a representative example of our experimental results, we found that mean absolute error is less than 8 mm and max absolute error is about 45 mm; the standard deviation is 7.42, 4.76, and 27.08 mm for, respectively, the x-, y- and z-component. The range of each component is $[100; 1000\text{ mm}]$ for x, $[-400; 1000\text{ mm}]$ for y and $[500; 4000\text{ mm}]$ for z.

4.3. Real data

Hereafter, we present some results achieved from real images. The image processing method is described in Ref. [20]. Let us recall the key points. The original coloured image is first converted into the CIE-Lab space. Within the L-image, a self-adaptive thresholding is performed, followed by a morphological squeletization, a Hough transform and the recovering of intersecting points. Within the ab-image, a process to determine the projected colours from the apparent ones is performed which permits to decode the pattern.

The structured light system is composed by an RGB camera, a computer and an electronic slide projector. The coloured pattern is shaped in a 512×512 RGB image which is projected on the measuring scene using the projector, and the scene is then captured by the camera into the computer memory. The reader is pointed to Refs. [4,20] to focus on pattern segmentation and decoding.

In the following, we go on to the reconstruction results, giving some qualitative results.

4.3.1. Scene I

The scene is composed by three geometric and achromatic objects illuminated by the coloured pattern, as shown in Fig. 8. We proceeded in two steps: first a projective reconstruction using the canonical representation, and then a Euclidean reconstruction by adding constraints obtained

Table 1
Errors on reconstruction with uniform noise ±1

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
Standard deviation			0.610	0.655	3.222

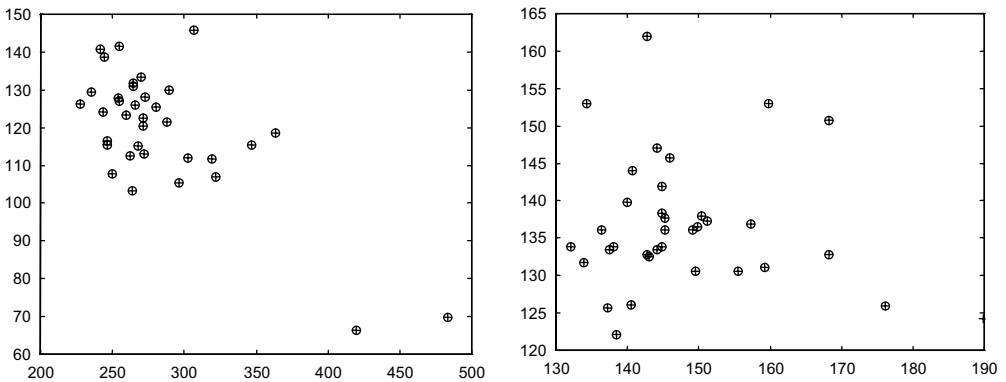

Fig. 7. Validation of the projective reconstruction, left and right image planes. Circles represent real image points and crosses represent projectively reconstructed ones.

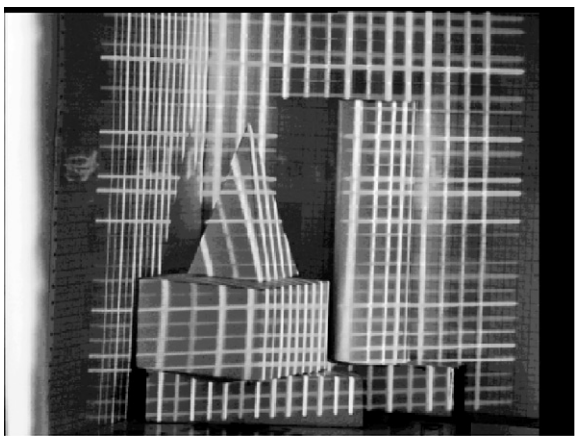

Fig. 8. Structured lighted image of the first scene.

from image analysis (note that, at the time, the constraints are generated manually).

In Fig. 9, we present the back-projection of the projective reconstruction onto the image plane and the projector plane obtained with the linear method (through $\mathbf{P}_{proj}$ and $\mathbf{P}'_{proj}$). Circles represent the real 2D points and crosses the back-projected ones. The projection matrices and the 3D projective points computed with the linear method are used as initializations for the iterative method: the results are clearly improved as shown in Fig. 10 and quantified in Table 2 (this time $\mathbf{P}_{proj}$ and $\mathbf{P}'_{proj}$ are given by the non-linear method). The maximum absolute 2D error is 3.069 pixels and the mean absolute 2D error is 0.204 pixels in the projector plane and 2.715 pixels and 0.169 pixels, respectively, in the image plane.

At this point, the projective reconstruction is validated (the point 2 in Section 3.5 is performed). We now have to

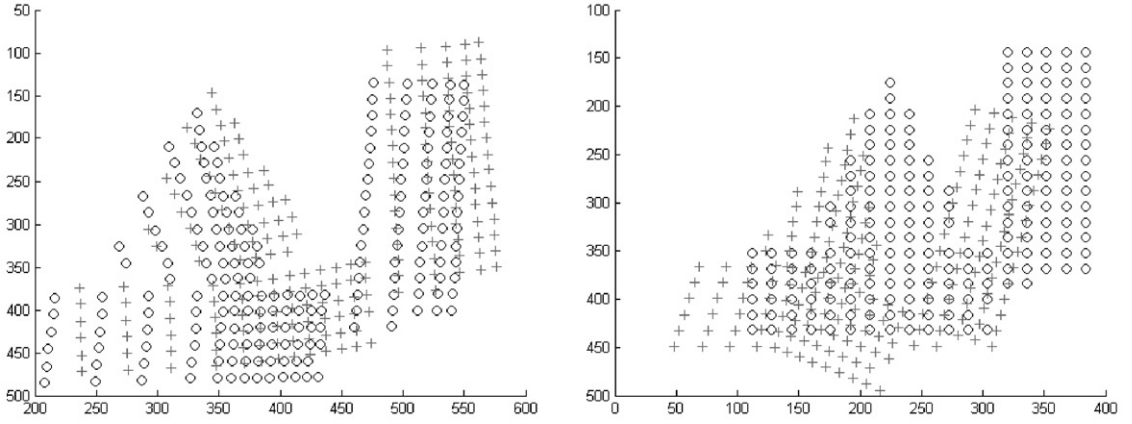

Fig. 9. Back-projection of the linear reconstruction method. Image plane (left) and projector plane (right).

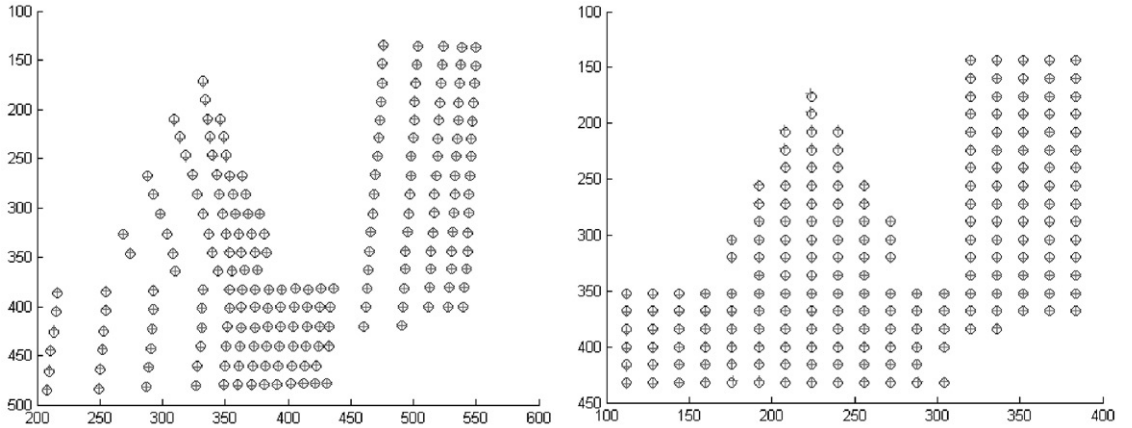

Fig. 10. Back-projection of the non-linear reconstruction method. Image plane (left) and projector plane (right).

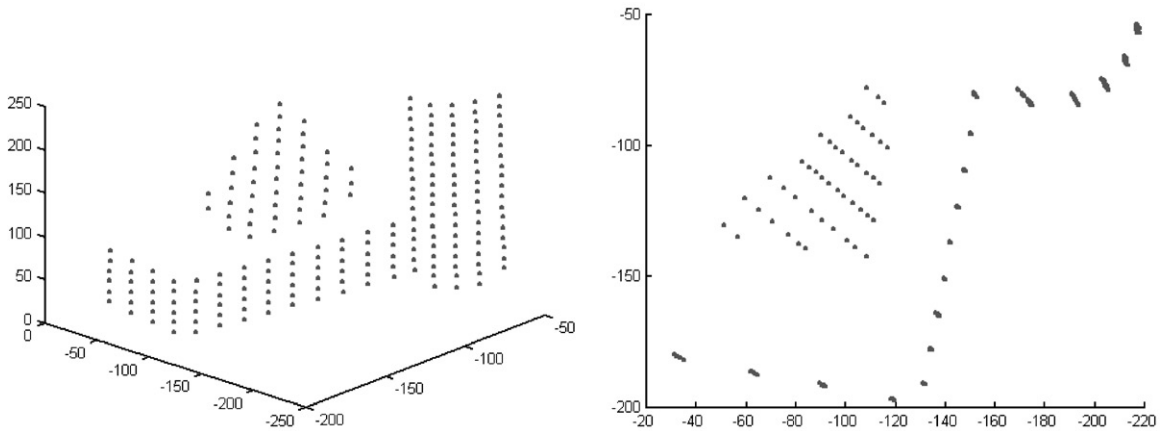

Fig. 11. Two views of the reconstructed scene (I).

Table 2
Residual 2D error

		Max. absolute error (pixels)	Mean absolute error (pixels)
Linear method	Camera	50.084	18.428
	Projector	88.143	32.751
Iterative method	Camera	3.069	0.204
	Projector	2.715	0.169

rectify it into the Euclidean space by following the steps of point 3 in Section 3.5: fix the origin, fix horizontal and vertical planes, generate parallelogram and orthogonality constraint, fix an arbitrary scale factor, etc. The results obtained are shown in Fig. 11. It can be seen that the three objects are globally well-reconstructed. Parallelism and orthogonality are recovered with a sufficient precision and proportions seem to be preserved.

4.3.2. Scene II

The goal of this experimentation is to validate the reconstruction method on a more realistic scene. The one shown in Fig. 12 is grabbed under real conditions of illumination; its size is about 1 m $\times$ 1 m. It is in an office environment, the image has been shot under the desk.

The highlighted lines of the figure have been reconstructed; the results are presented in Fig. 13. An arbitrary metric has been assigned, parallelogram and orthogonality constraints have been generated. Vertices of highlighted polygons are the reconstructed points and the lines which compose it show the geometrical constraints (parallel lines give parallelism constraints, orthogonal lines give orthogonality constraints, etc.)

Similarly to the previous scene, it can be noticed that parallelism and orthogonality are satisfactorily reconstructed, as well as the image proportion.

5. Conclusions

This article presents a method to perform Euclidean reconstruction from an uncalibrated structured light sensor independently of the scene geometry, by assuming that the sensor behaviour is affine or that it can be approximated by an affine camera model. 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 world reference frame that it has been chosen. 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 elements of the collineation matrix. Besides, we show that projecting a known grid pattern of light onto the scene permits to retrieve intrinsic geometrical knowledge about this scene as parallelism and orthogonality. The major

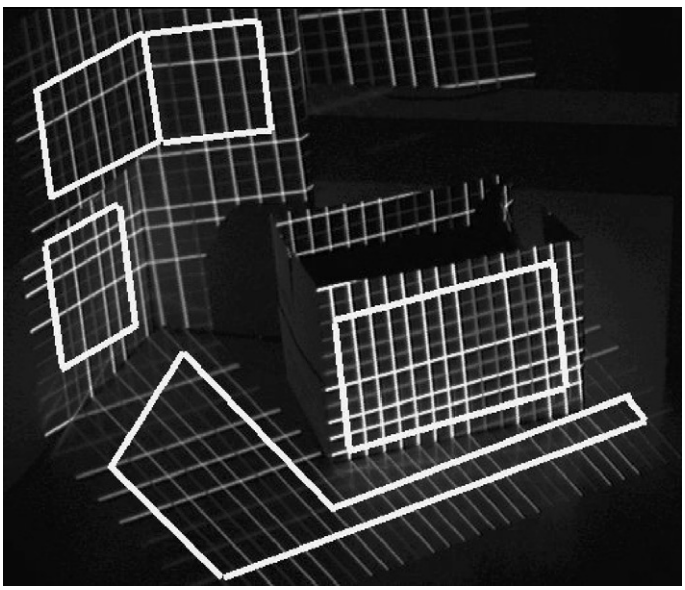

Fig. 12. Structured lighted image of the second scene.

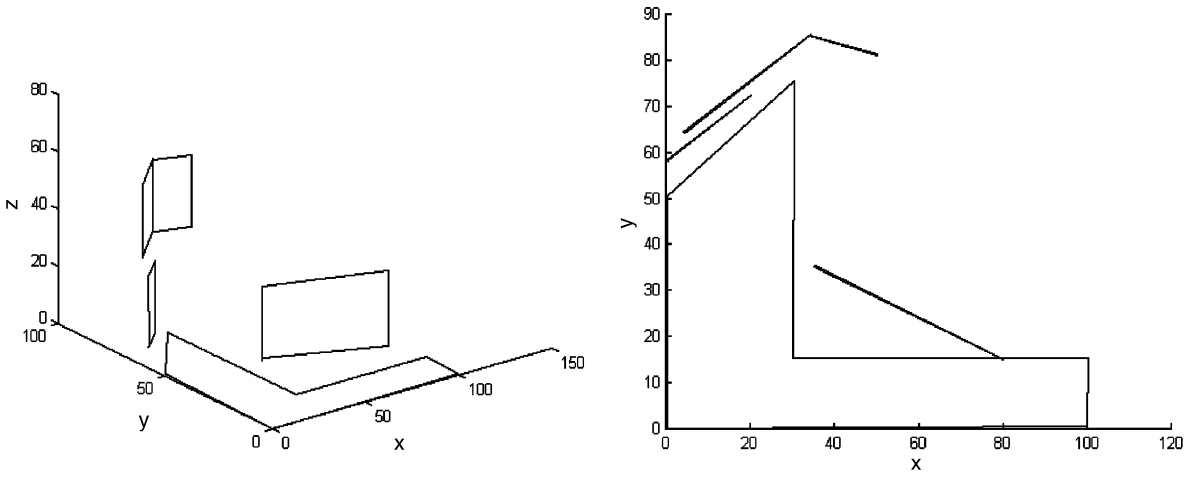

Fig. 13. Two views of the reconstructed scene (II).

contribution of the paper is to show that structured light can be used to deduce geometrical constraints of the scene, which are used to reconstruct the scene without any previous calibration. As no constraint is required on the projection matrices, this approach allows us to reconstruct without considering potential change of 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. However, the automation of the whole process is necessary. 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. As a further work, we intend to automate the constraints generation, that is, to formulate them mathematically and to solve them from the image segmentation and decoding.

6. Summary

This paper deals with uncalibrated reconstruction through structured lighting. In a structured light system, unlike classical stereovision, the second camera is replaced by a light source that projects a known pattern of light onto the scene. The main goal of this work is to provide adaptive capabilities to this kind of sensor which allows to use it in mobile robotics or dynamic visual inspection. First, we present a survey of the most relevant techniques of uncalibrated reconstruction. It is shown that, due to the fact that any movement of the light source produces a movement of the pattern (i.e. of the 3D points), the reconstruction has to be performed from a single camera shot and a single pattern projection. Thus, we first focus on a projective reconstruction method based on the canonical representation of views, which requires only pixel correspondences, one view and one pattern projection. The reconstruction is performed in a projective frame, up to a projective transformation.

An Euclidean reconstruction can be recovered from a projective one since Euclidean transformations are a sub-group of projective transformations. In other words, there exists a collineation matrix which permits to pass from projective to Euclidean. This matrix can be assessed by constraining its entries with geometrical knowledge grabbed from the scene. We describe how the pattern projection is used to acquire geometrical knowledge as parallelism, orthogonality, horizontality and verticality. Moreover, 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, performed both on simulated and real data, are presented and discussed.

References

[1] O. Faugeras, Three-Dimensional Computer Vision: A Geometric Viewpoint, MIT Press, Cambridge, MA, 1993.

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

[3] J. Batlle, E. Mouaddib, J. Salvi, Recent progress in coded structured light as a technique to solve the correspondence problem. A survey, Pattern Recognition 31 (7) (1998) 963–982.

[4] J. Salvi, J. Batlle, E. Mouaddib, A robust-coded pattern projection for dynamic measurement of moving scenes, Pattern Recognition Lett. 19 (1998) 1055–1065.
[5] M. Proesmans, L. Van Gool, A. Oosterlinck, One-shot active range acquisition, Proceedings of the International Conference on Pattern Recognition, Vienna, Austria, 1996, pp. 336–340.
[6] J.M. Sotoca, M. Buendia, J.M. Iñesta, A new structured light calibration method for large surface topography, Pattern Recognition Appl. Frontiers Artif. Intell. 56 (2000) 261–270.
[7] D.Q. Huynh, R.A. Owens, P.E. Hartman, Calibrating a structured light stripe system: a novel approach, Int. J. Comput. Vision 33 (1) (1999) 73–86.
[8] J.J Koenderink, A.J. Van Doorn, Affine structure from motion, Technical Report, Utrecht University, Utrecht, Netherlands, October 1989.
[9] R. Hartley, R. Gupta, T. Chang, Stereo from uncalibrated cameras, Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA, 1992, pp. 761–764.
[10] R. Mohr, B. Boufama, P. Brand, Accurate projective reconstruction, Proceedings of the Second ESPRIT-ARPA-NSF Workshop on Invariance, Azores, 1993, pp. 257–276.
[11] O. Faugeras, Q.-T. Luong, S.J. Maybank, Camera self-calibration: theory and experiments, Proceedings of the Second European Conference on Computer Vision, Santa Maria Ligure, Italia, 1992, pp. 321–334.
[12] R.I. Hartley, Kruppa's equations derived from the fundamental matrix, IEEE Trans. Pattern Anal. Mach. Intell. 19 (2) (1997) 133–135.
[13] R. Hartley, Euclidean reconstruction from uncalibrated views, Lecture Notes in Computer Science, Vol. 825, 1994, pp. 237–256.
[14] A. Heyden, K. Åström, Euclidean reconstruction from constant intrinsic parameters, Proceedings of the International Conference on Pattern Recognition, Vienna, Austria, 1996, pp. 339–343.
[15] B. Boufama, R. Mohr, F. Veillon, Euclidean constraints for uncalibrated reconstruction, Proceedings of the Fourth International Conference on Computer Vision, Berlin, Germany, 1993, pp. 466–470.
[16] Z. Zhang, K. Isono, S. Akamatsu, Euclidean structure from uncalibrated images using fuzzy domain knowledge: application to facial images synthesis, Proceedings of the International Conference on Computer Vision, Bombay, India, 1998, pp. 784–789.
[17] Q.-T. Luong, T. Viéville, Canonic representations for the geometries of multiple projective views, Proceedings of the Third European Conference on Computer Vision, Stockholm, Sweden, 1994, pp. 589–599.
[18] D.W. Marquardt, An algorithm for the estimation of nonlinear parameters, Soc. Ind. Appl. Math. J. 11 (1963) 431–441.
[19] K.E. Aström, L. Morin, Random cross ratios, Proceedings of the Ninth Scandinavian Conference on Image Analysis, Uppsala, Sweden, 1995, pp. 1053–1061.
[20] D. Fofi, Navigation d'un véhicule intelligent à l'aide d'un capteur de vision en lumière structurée et codée, Ph.D. Thesis, CREA, Université de Picardie Jules Verne, 2001.

About the Author—DAVID FOFI received the M.Sc. degree in Computer Science (computer vision, signal and image processing) in the University of Cergy-Pontoise, ENSEA, in 1997. Since 1998, he is involved in the study of structured light vision for mobile robotics. In September 2001, he received the Ph.D. degree in Robotics and Computer Vision from the University of Picardie Jules Verne. Since September 2002, he works as an associate professor in the Le2i—IUT Le Creusot.

About the Author—JOAQUIM SALVI received the B.Sc. degree in Computer Science in the Polytechnic University of Catalunya in 1993. He received the M.Sc. degree in Computer Science in 1996 and the Ph.D. degree in Electrical Engineering in 1998 both from the University of Girona. At present he is an associate professor in the Electronics, Computer Science and Automation Department of University of Girona where he is involved in research projects on mobile robotics and computer vision. His current interests are in the field of stereovision, structured light and scene reconstruction.

About the Author—EL MUSTAPHA MOUADDIB received the Ph.D. degree in robotics at the University of Picardie Jules Verne, Amiens, France, in 1991 and HDR in December 1999. He is professor at the same University where he teaches Robotics and Computer Vision. His main interests are computer vision, omnidirectional vision, mobile robotics and perception. Since 2001, he is the head of the "Centre de Robotique, d'Electrotechnique et d'Automatique", where he is involved in research projects on robot navigation with a conic sensor and in particular localisation and obstacle detection.