HAL
open science

Relative 3D Reconstruction Using Multiple Uncalibrated Images

Roger Mohr, Long Quan, Françoise Veillon, Boubakeur Boufama

To cite this version:

Roger Mohr, Long Quan, Françoise Veillon, Boubakeur Boufama. Relative 3D Reconstruction Using Multiple Uncalibrated Images. [Technical Report] RT 84-IMAG- 12 LIFIA, 1992, pp.25. inria-00548434

HAL Id: inria-00548434

https://inria.hal.science/inria-00548434

Submitted on 31 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LIFIA
INSTITUT IMAG
46, AVENUE FELIX VIALLET
38031 GRENOBLE CEDEX
FRANCE

Centre National de la Recherche Scientifique, URA 394 - Institut National Polytechnique de Grenoble

RAPPORT TECHNIQUE

Relative 3D reconstruction using multiple uncalibrated images

R. Mohr, L. Quan, F. Veillon & B. Boufama

RT 84-IMAG-12 LIFIA Juin 1992

Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle

Téléphone (33) 76 57 45 00

Relative 3D Reconstruction Using Multiple Uncalibrated Images

R. Mohr L. Quan F. Veillon B. Boufama

(RT 84-IMAG 12-LIFIA 1992)

LIFIA - IRIMAG,
46, avenue Felix Viallet,
38031 Grenoble, France

Abstract

In this paper, we show how relative 3D reconstruction from multiple uncalibrated images can be achieved through reference points. The original contributions with respect to other related works in the field are mainly a direct non linear method for relative 3D reconstruction, a geometrical method to select the set of reference points among all image points and a geometrical interpretation of the linear reconstruction method. Experimental results from both simulated and real image sequences are presented.

1 Relative positioning

From a single image, no depth can be computed without *a priori* information. Even more, no invariant can be computed from a general set of points [3]. This problem becomes feasible using multiple images. The process is composed of two major steps. First image features are matched in the different images. Then, from such a correspondence, depth is easily computed using standard triangulation. This kind of classical technique needs careful calibration of the imaging system and usually it is performed by computing each camera parameters in an aboslute reference frame.

This approach suffers from several drawbacks: firstly the calibration process is an error sensitive process, secondly it cannot always be performed off line, particularly when the imaging system is obtained by a dynamic

system with zooming, focusing and moving. Similarly stereo vision with a moving camera is impossible as the standard tool for locating the position of a camera with translation and rotation does not reach the required precision for calibrating such a multistereo system. Introducing in each image beacons with exact known position may overcome these drawbacks: calibration and reconstruction are then solved in the same process [2, 1]. But for many problems it is impossible to provide such carefully positioned reference points.

The alternative approach is to use points in the scene as reference frame without knowing their coordinates nor the camera parameters. This has been investigated by several researchers these past few years. It is also the goal of this paper.

Using parallel projection for the imaging system, Kœnderink and van Doorn [10] were able to choose 4 points as an affine reference frame in the scene and reconstruct all other points seen in at least two images. Using also parallel projection but with a very different approach, Tomasi and Kanade [17] were able to reconstruct a scene from images. Dealing with real images, their results were satisfactory as the imaging system was equipped with a long focal lens. Their contribution is also a nice way to deal with erroneous data using SVD[1] for filtering noise from information.

Approaching the problem for real perspective projection implies to leave the affine geometry for the projective geometry. Two major orientations were developed these two last years. Sparr [16, 15] developed an affine shape descriptor; such a descriptor contains the affine information of the relative position of this group of points; it is related to its perspective projection by a major theorem which allows to recover relative depth or shape. This approach is particularly well suited when affine information can be used like observing parallelograms, of when the imaging system has a known affine reference frame.

Using the more standard tool of projective geometry, we proposed to generalize the Kœnderink and Doorn's method with additional reference points [11, 13]. The right way to approach the problem was first described by Faugeras [4]. This paper is largely inspired by his paper and uses the epipolar geometry reconstruction method he provides.

The original contributions of this paper are a geometrical way to choose among the set of points those which can be selected as reference points, a geometrical interpretation of the reconstruction, a direct solution of the

[1]Singular Value Decomposition

system of equations and a discussion around experimental results.

First the paper describes how reference points in the scene provide us a way to reconstruct the scene, and why this solution can only be defined up to a projective transformation, i.e. a collineation. Then Faugeras' method for computing the epipolar geometry is presented. From the epipolar geometry it is shown how to solve the basic equations. Section 3 provides basic results on the location of the projection of coplanar points. This result allows to provide a geometrical interpretation of the relative reconstruction, as Kœnderink [10] did it in the affine case. It allows also to derive a computational way to choose among the points present in the scene, those which can be selected for the reference frame. Finally we describe an implementation which allows to solve the problem in presence of noise using redundant data. This is done by an implementation of the parameters estimation theory using Levenberg-Marquardt algorithm. Robustness of this implementation is discussed from the results obtained on both synthetic and real data.

Two basic assumptions are made here. First we assume that the reader is familiar with elementary projective geometry, as it can be found in the first chapters of [14] (see also [5]). We also assume throughout this paper that the imaging system is a perfect perspective projection, i.e. the camera is a perfect pinhole. However this point will be discussed with the interpretation of the experimental results.

2 Using scene references points

This section provides the basic equations of 3D reconstruction problem, together with the self calibration problem. This derivation was developed independently from these recently published by Faugeras in [4]. The basic starting point is similar to this work, however the way to solve it was influenced by the way photogrametrists simutaneously calibrate their camera and reconstruct the scene, by using carefully located beacons (cf. [1]).

We consider m views of a scene ($m \geq 2$); it is assumed that n points have been matched in all the images, thus providing $n \times m$ image points. The assumption that the scene points appear in all the images is not essential but simplifies the explanation here.

$\{M_i, i = 1, \ldots, n\}$ is the (unknown) set of 3D points projected in each image, represented by a column vector of its four yet unknown homogeneous coordinates.

2.1 The basic equations

For each image j, the point M_i, represented by a column vector of its homogeneous coordinates $(x_i, y_i, z_i, t_i)^T$ or its usual non homogeneous coordinates $(X_i, Y_i, Z_i)^T = (\frac{x_i}{t_i}, \frac{y_i}{t_i}, \frac{z_i}{t_i})^T$, is projected as the point $\mathbf{m}_{ij}$, represented by a column vector of its three homogeneous coordinates $(u_{ij}, v_{ij}, w_{ij})^T$ or its usual non homogeneous coordinates $(x_{ij}, y_{ij})^T$. Let $\mathbf{P}_j$ be the 3×4 projection matrix of the jth camera.

We have for homogeneous coordinates

$$\rho_{ij}\mathbf{m}_{ij} = \mathbf{P}_j\mathbf{M}_i, i = 1, \ldots, n, j = 1, \ldots, m \tag{1}$$

where ρ_{ij} is an unknown scaling factor (different for each image point).

Equation 1 is usually written in the following way, hiding the scaling factor, using the non homogeneous coordinates of the image points:

$$x_{ij} = \frac{m_{11}^{(j)}x_i + m_{12}^{(j)}y_i + m_{13}^{(j)}z_i + m_{14}^{(j)}t_i}{m_{31}^{(j)}x_i + m_{32}^{(j)}y_i + m_{33}^{(j)}z_i + m_{34}^{(j)}t_i} \tag{2}$$

$$y_{ij} = \frac{m_{21}^{(j)}x_i + m_{22}^{(j)}y_i + m_{23}^{(j)}z_i + m_{24}^{(j)}t_i}{m_{31}^{(j)}x_i + m_{32}^{(j)}y_i + m_{33}^{(j)}z_i + m_{34}^{(j)}t_i} \tag{3}$$

These equations express nothing else than the collinearity of the space points and their corresponding projection points.

As we have n points and m images, this leads us to $2 \times n \times m$ equations. The unknowns are $11 \times m$ for the P_j which are defined up to a scaling factor, plus $3 \times n$ for the M_i. So if m and n are large enough we have a redondant set of equations.

It is easy to understand that the solution for the equation 1 is not unique. For instance, if the origin is translated, all the coordinates will be translated and this will induce new matrices P_j satisfying 1. More generally, let A be a spatial collineation represented by its 4×4 invertible matrix. If $P_j, j = 1, \ldots, m$ and $M_i, i = 1, \ldots, n$ are a solution to 1, so are obviously P_jA^{-1} and AM_i, as

$$\rho_{ij}\mathbf{m}_{ij} = (\mathbf{P}_j\mathbf{A}^{-1})(\mathbf{A}\mathbf{M}_i), i = 1, \ldots, n, j = 1, \ldots, m$$

Therefore is established the first result:

Theorem: *the solution of the system 1 can only be defined up to a collineation.*

As a consequence of this result, a basis for any 3D collineation can be arbitraryly chosen in the 3D space. For a projective space $I\!P^3$, 5 algebraically free points form a basis, that is a set of 5 points, no four of them coplanar. We will come back to how to choose for such a basis later in 3.1. For convenience, we assume here that the first five points M_i can be chosen to form such a basis; their coordinates can be assigned to the canonical ones:

$$(1,0,0,0)^T, (0,1,0,0)^T, (0,0,1,0)^T, (0,0,0,1) \text{ and } (1,1,1,1)^T$$

The remaining part of this section is devoted to the problem of building from these now fixed reference points an explicit solution.

2.2 Direct non linear reconstruction

From the above section, the most direct way is to try to solve this system of non linear equations. As the projective coordinates of the spatial points are defined up to a constant, so for each point, the constraint $x_i^2+y_i^2+z_i^2+t_i^2 = 1$ can be added. Since the system is an overdetermined one, we can hope to solve it by standard least squares technique. The problem can be formulated as minimizing over

$$(x_i, y_i, z_i, t_i, m_{11}^{(j)}, ...m_{34}^{(j)}) \text{ for } i = 1, \ldots, m, j = 1, \ldots, n;$$

$$\sum_{k=1}^{2\times m\times n+n} \left(\frac{f_k(x_{ij}, y_{ij}; x_i, y_i, z_i, t_i, m_{11}^{(j)}, ...m_{34}^{(j)})}{\sigma_k}\right)^2 \tag{4}$$

where $f_k(\cdot)$ is either

$$x_{ij} - \frac{m_{11}^{(j)}x_i + m_{12}^{(j)}y_i + m_{13}^{(j)}z_i + m_{14}^{(j)}t_i}{m_{31}^{(j)}x_i + m_{32}^{(j)}y_i + m_{33}^{(j)}z_i + m_{34}^{(j)}t_i}$$

or

$$y_{ij} - \frac{m_{21}^{(j)}x_i + m_{22}^{(j)}y_i + m_{23}^{(j)}z_i + m_{24}^{(j)}t_i}{m_{31}^{(j)}x_i + m_{32}^{(j)}y_i + m_{33}^{(j)}z_i + m_{34}^{(j)}t_i}$$

or

$$x_i^2 + y_i^2 + z_i^2 + t_i^2 - 1$$

σ_k is the standard deviation of each image measure, x_{ij} or y_{ij}. On the other hand, it can also be considered as the weight for each function. So the problem is a general weighted least squares estimation. The only known measures are the image points (x_{ij}, y_{ij}). All others are unknown parameters to estimate. In addition, as each projection matrix is equally defined up to a constant, we can for example impose $m_{34}^{(j)} = 1$.

This can be solved by the standard nonlinear least squares algorithm due to Levenberg-Marquardt [12]. It has to be mentioned that this system leads to $m + 2 \times n \times m$ equations in $11 \times m + 4 \times n$ unknowns, which is quite large. More technical discussions about it will be found in section 4.

2.3 Linear reconstruction by Faugeras

Faugeras presented in [5] an elegant linear reconstruction method with the tricky use of the epipolar geometry. The basic idea is that he first tries to determine the projection matrix up to a collineation. Once five general points are selected as the reference points and are assigned the standard basis projective coordinates, this is equivalent to have 5 image points and space points matches. Thus in each projection matrix remains only one unknown parameter. Now the author supposes that the epipolar geometry has been established, for example by the matching of at least 8 points. Since the epipoles are intrinsically related to the projection matrix, thanks to the known epipoles, each projection matrix is entirely determined up to a collineation (or determined projectively). Once each projection matrix is known, the reconstruction becomes straightforward resolution of a linear equations set.

We can note that Faugeras still performs a planar collineation to obtain the projection matrix in a simpler form. This basis change in image plane is not essential. It only makes the projective matrix to be simpler. And also it is important to note that this linear reconstruction is essentially based on the previous establishment of epipolar geometry, in the absence of which the linear reconstruction is not possible. We will have a more detailed discussion below.

2.4 Solving the epipolar geometry

The epipolar geometry plays the most fundamental role for motion analysis. Knowing this geometry is equivalent to knowing the relative orientation of two cameras. This consideration has been largely discussed in [7]. With

Sturm's algorithm, the epipolar geometry is determined with seven points, the solutions are obtained from the intersection points of cubics curves. This approach becomes quickly prohibitive when real image points are concerned. However, recently, Faugeras *et al.* pointed out that this geometry can be computed in a linear way when at least 8 point matches between the two images are available. The equation is formally the same as that has been proposed by Longuet-Higgins a few years ago, well known as 8-point algorithm, but this equation admits a nice projective interpretation. So this section is largely inspired from [4]. We will get the same equation. We rephrase the demonstration here, but with some highlights on geometrical and algebraic interpretation of the equations.

Let $\mathbf{m} = (x, y, t)$ be a point in the first image and let $\mathbf{e} = (u, v, w)$ be the epipole point with respect to image 2. The three homogeneous coordinates (a, b, c) of the epipolar line $\mathbf{l}$ going through $\mathbf{e}$ and $\mathbf{m}$ are $\mathbf{m} \times \mathbf{e}$ where $\times$ denotes the cross product: obviously $\mathbf{l} \cdot \mathbf{m}^T = \mathbf{l} \cdot \mathbf{e}^T = 0$. The mapping $\mathbf{m} \longrightarrow \mathbf{m} \times \mathbf{e}$ is linear and can be represented by a skew symmetric matrix $\mathbf{C}$ of rank 2:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} yw - zv \\ zu - xw \\ xv - yu \end{pmatrix} = \begin{pmatrix} 0 & w & -z \\ -w & 0 & u \\ z & -u & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{5}$$

The mapping of each epipolar line $\mathbf{l}$ from image 1 to its corresponding epipolar line $\mathbf{l}'$ in image 2 is a collineation defined in the dual space of lines in $\mathcal{P}^2$. Let $\mathbf{A}$ be one such collineation: $\mathbf{l}'^T = \mathbf{A}\mathbf{l}^T$.

It is defined by the correspondence of 3 distinct epipolar lines. The first two correspondences provide four constraints as the degree of freedom of a line is 2. As the third line in correspondence belongs to the pencil defined by the two first ones, the third correspondence only adds one more constraint. So $\mathbf{A}$ only has five constraints for eight degrees of freedom.

Let $\mathbf{E} = \mathbf{A}\mathbf{C}$. Using (5) we get

$$\mathbf{l}'^T = \mathbf{A}\mathbf{C}\mathbf{m}^T \tag{6}$$

As $\mathbf{A}$ has rank 3 and $\mathbf{C}$ has rank 2, $\mathbf{E}$ has rank 2. As the kernel of $\mathbf{C}$ is obviously $\lambda\mathbf{e} \equiv \mathbf{e}$, the epipole is the kernel of $\mathbf{E}$.

Let now $\mathbf{m}'$ be the corresponding point of $\mathbf{m}$. Using 5 the epipolar constraint $\mathbf{m}'^T \cdot \mathbf{l}' = 0$ can be rewritten $\mathbf{m}'^T\mathbf{E}\mathbf{m} = 0$. So each matching between the two images provides a constraint on $\mathbf{E}$, and as $\mathbf{E}$ is defined up to a scaling factor. Thus 8 independant constraints will allow us to compute

linearly E and therefore to get the epipolar geometry. Notice that E is determined by 7 parameters: C has 2 (the epipole position) and A has 5. But allowing a redondant set of constraints provides a unique solution which can be linearly computed. Faugeras *et al.* [6] has pointed out that for an accurate estimation of E, a non linear technique has to be used. In fact, the epipolar geometry provides only one point to line transformation in $\mathcal{P}^2$. This is known as a *correlation*, also a linear transformation, which, instead of collineation transforming points to points, systematically dualizes it. The equation derived above is just the concept of conjugacy of a pair of points with respect to the correlation: a point p is said to be conjugate with respect to E to a point p' if p' lies on the line $\mathbf{l}' = \mathbf{E}\mathbf{p}$ from the point p, thus we obtain the basic bilinear equation $\mathbf{x}'^T\mathbf{E}\mathbf{x} = 0$. Get transposed the equation, we get the other linear transformation from the second image plane to the first image plane which is the transposed E. As in our case all transformed lines form a pencil going through the epipole, so directly the rank of this line space is at most 2 and in this case the kernel of this transformation is the epipole.

This computation can be efficiently done via SVD. Firstly, using SVD, the system of homogeneous equations can be solved in the least squares sense. As our way to solve the system 4 does not require the epipolar geometry, we compute E linearly and efficiently. The solution obtained this way is less accurate than that computed by Faugeras *et al.* [6]; it is however sufficient to get epipolar lines for checking the matches. We just show in Figures 1 an example of the obtained epipolar geometry with the tracked points (discussed later in section 4).

3 Geometrical reconstruction

In this section, we will show some very interesting geometric properties once the epipolar geometry has been established. In particular, we can determine if any fourth point is coplanar with the plane defined by any three other points, of course only through operation in the image planes. That leads to an automatic selection of general reference points from image planes and point reconstruction in a geometric way, thus providing a geometric interpretation of the linear reconstruction method. We assume through this section that the epipolar geometry has been established using for instance the method described in section 2.4.

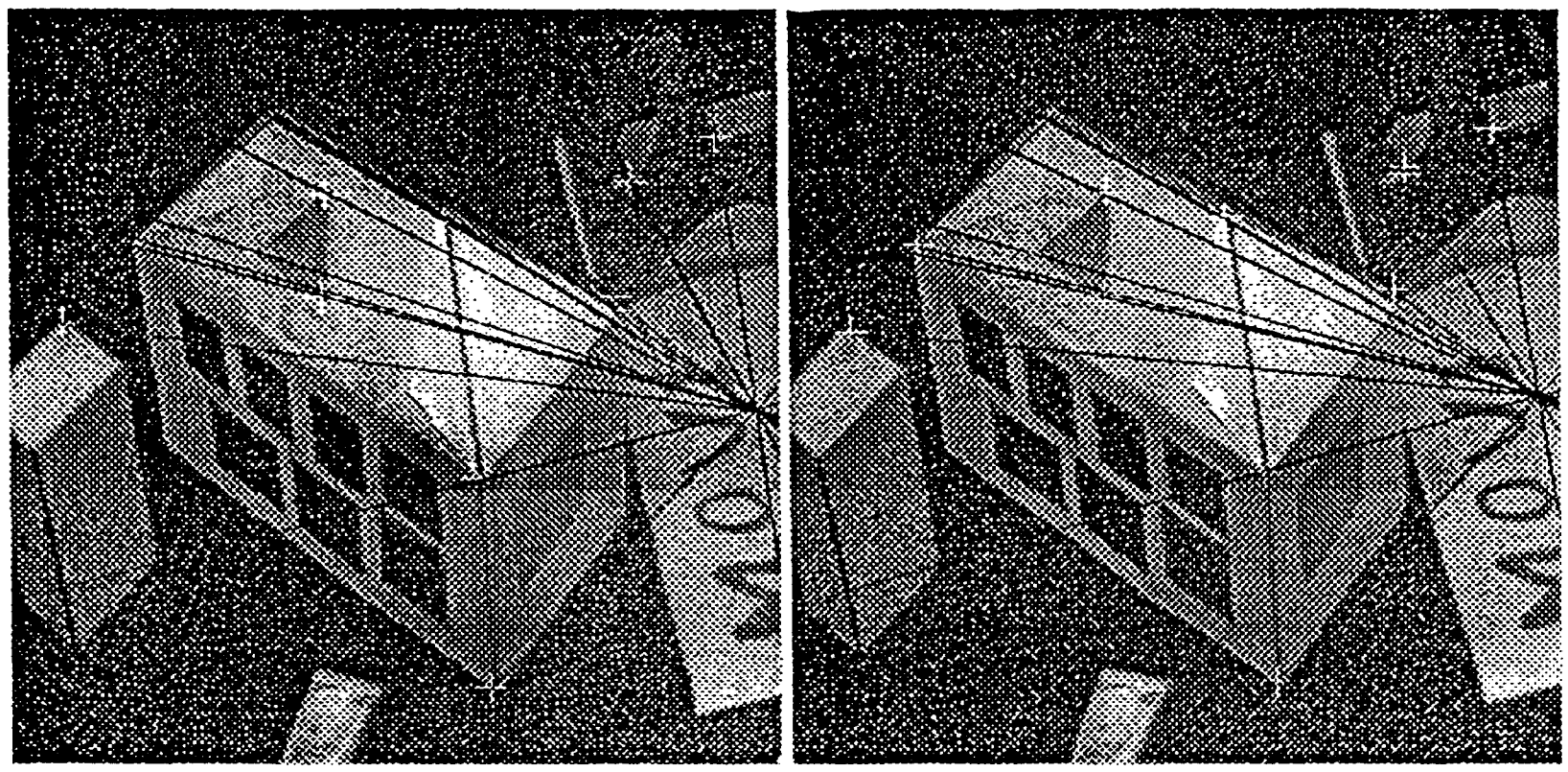

Figure 1: In this pair of images, the tracked points are marked by a cross and the epipolar line of each point is drawn by a dark line.

3.1 The coplanarity test

As we assume here that the epipolar constraint is known, we know the essential matrix E which contains all this information [4, 9]. $\mathbf{E}$ is a 3×3 matrix such that from the point $\mathbf{m} = (x, y, t)^T$ in image 1, the corresponding epipolar line l' in image 2 has its coefficients satisfying $\mathbf{l}' = (a', b', c')^T = \mathbf{Em}$.

Now, consider figure 2. It displays two images of four 3D points A, B, C, D, projected in the two images. The dashed lines correspond to some of the epipolar lines going through each of the vertices of the quadrangles. The epipolar constraint specifies that the epipolar line corresponding to c passes through c', and conversely.

If A, B, C, D are coplanar, then the diagonals intersect in this 3D space plane in a point M which is projected respectively as m and m'. Therefore m and m' have to satisfy the epipolar constraint too, as it is displayed in Fig. 2.

Conversely consider the case where A, B, C, D are not coplanar. The diagonals are no more in the same plane and therefore does not intersect in the space. So m is the image of two 3D points, M_1 lying on (AC), and N_1 lying on (BD). Similarly m' is the image of M_2 and N_2. If the central point O' of the second image is not in the plane defined by (ACO), nor in the

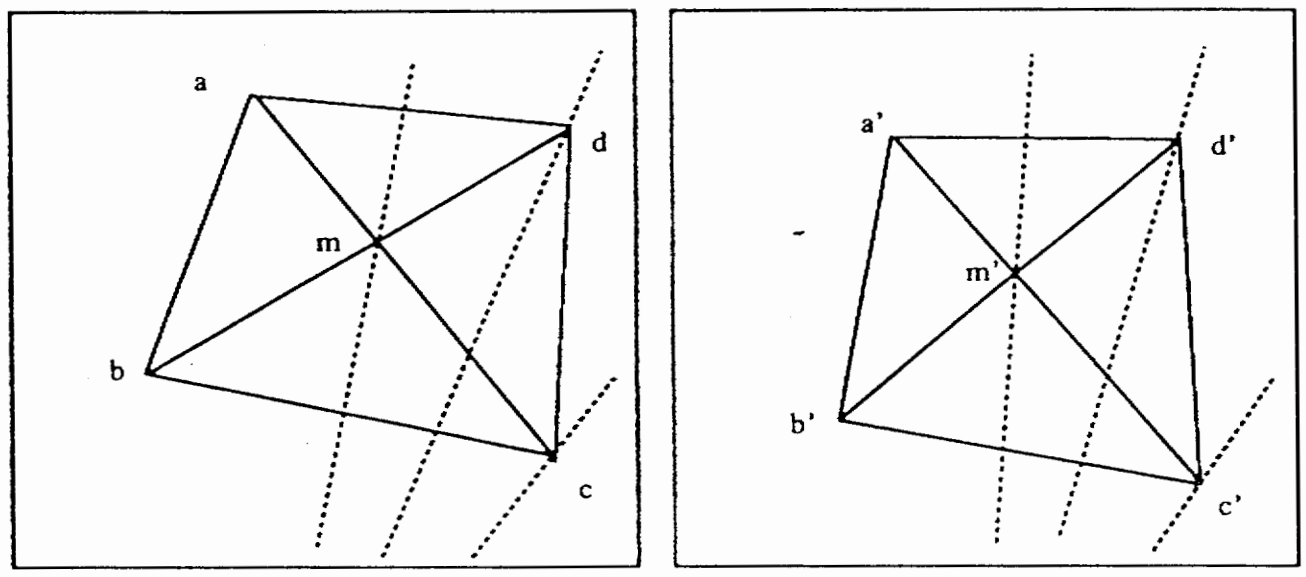

Figure 2: Match of diagonal intersections with epipolar constraint

plane (BDO), then the 2 view lines (Om) and $(O'm')$ does not intersect, and therefore the points m and m' are not in epipolar correspondence.

The condition that O' does not lie in the plane (OAC) is equivalent to the condition that the epipole in the first image does not lay on (ac), which is therefore checked easily. Notice that in such a case, we can choose as diagonals (AB) and (CD) instead of (AC) and (BD). Therefore the only condition we reach for applying this method is to have none of the projections a, b, c, d being the epipole.

So we proved that

Theorem: *If neihter a, b, c, nor d are the epipole point of image 2 with respect of image 1, then it exists at least one diagonal intersection m such that m and its corresponding intersection m' satisfy the epipolar constraint if and only if A, B, C, D are coplanar.*

We only had a look on the geometrical problem here. Technical problems like finding the points which are as less coplanar as possible and widely spread in the field of view are not addressed here, but can be easily deduced from the geometrical results.

In fact this theorem leads to a usefull and straighforward construction technique. Observing three points and a line in an image, it is possible to reconstruct the intersection of this line with the plane defined by the three points.

Let a, b, c be the projections of A, B, C in image 1, and let l be the projection of the line L. Let d be a point on l; the image coordinates of d depend linearly on a single parameter t and will be denoted as $d(t)$. Its corresponding projection in image 2, $d'(t)$, is the intersection of l' and

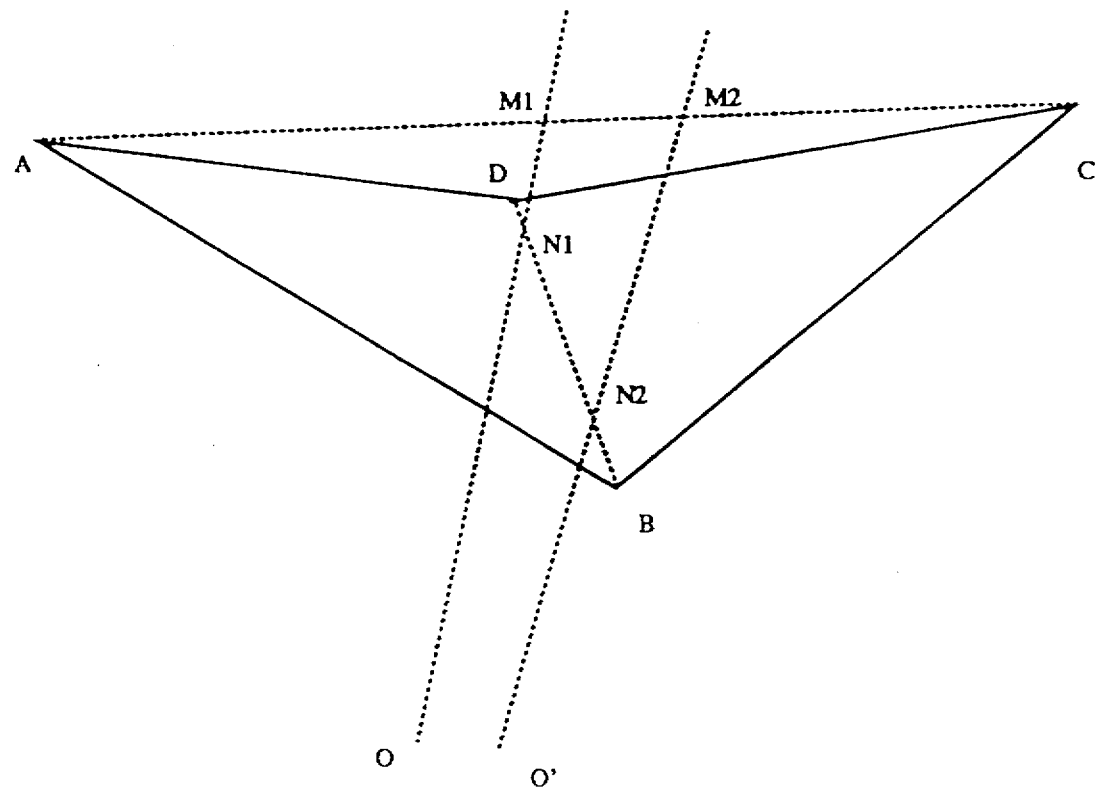

Figure 3: Four non coplanar points in space

the epipolar line associated with d. So $d'(t)$ also depends linearly on t. Finding m and m' as intersection of the diagonals provides coordinates of these two points still expressed linearly with respect of t. Stating now that $\mathbf{m}\mathbf{E}\mathbf{m}'^T = 0$ then leads to a second degree equation in t. Obviously one solution is that l and l' are epipolar line, so the only non trivial remaining solution is the one we are seeking for: the value of t for which $d(t)$ is the image of the intersection of L with the plane (A, B, C).

This leads to

Corollary : *the intersection of a line with a plane defined by three points can be algebraically computed from the image of these data, provided the essential matrix.*

Such a technical result is particularly useful for computing construction directly in the image without going through the 3D reconstruction. It allows for instance to compute several invariants for stereo images (*cf.* [8]).

3.2 Search for a 5 point basis

The above result can be directly applied to automatically select the necessary reference points from image points for projective reconstruction without any *a priori* spatial knowledge. Basically, we will be able to get rid of the coplanar reference points in the step 3 with the previous section's results.

Such a greedy algorithm could be:

1. choose any point for M_1 and M_2
2. choose for M_3 any point not aligned with M_1M_2
3. choose for M_4 any point such that it is not coplanar with $M_1M_2M_3$
4. choose for M_5 any point such that it is not coplanar with any face of the tetrahedron M_1, M_2, M_3, M_4.

This algorithm will give us a mathematically correct reference points set. In practice, reference points selection has also to take into account the precision of the measure in the image. It's better to take reference points as far as possible from each other. In this case, one improved version of the algorithm can be

1. choose for M_1 and M_2 the farthest points pair in one of the image.
2. choose for M_3 the farthest point from M_1M_2.
3. sort the other points according to the *distances* to the plane determined by the triangle $M_1M_2M_3$, choose for M_4 the one which has the maximum distance. The *distance* is not the orthogonal distance from the point to the plane as we expect (not possible at this step), it is the projection on the second image of the distance from the point to the plane along the first viewing line of that point (see Figure 4).
4. Sort the remaining points according to the maximum distance to any face of the tetrahedron M_1, M_2, M_3, M_4, choose for M_5 the point which has the maximum distance.

This improved version of reference points selection will provide us with a reasonably scattered points set.

3.3 Geometrical solution for point position

In this section, we are providing a geometric solution for relative projective shape reconstruction. To get the direct geometric reconstruction of the projective coordinates representing the projective shape, we start from the geometric definition of projective coordinates (cf. Figure 5). Take four reference points M_1, M_2, M_3 and M_4 as the vertices of the tetrahedron of

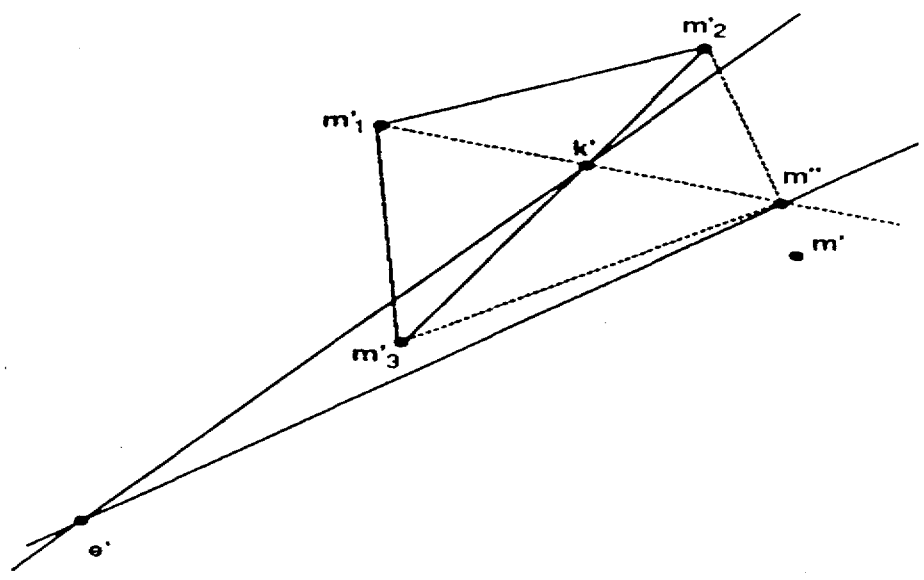

Figure 4: The *distance* is defined as that between m' and m''.

reference, and the fifth point M_5 as the unit point. For any point P whose homogeneous coordinates are $(x, y, z, t)^T$, we have one of the non homogeneous coordinates defined as

$$\frac{x}{t} = \{M_1, M_2; B, R\}$$

B and R are the intersection points of M_3A and M_3Q with the line M_1M_2 and in their turn A and Q are the intersection points of M_4M_5 and M_4P with the face $M_1M_2M_3$ of the tetrahedron of reference. $\frac{y}{t}$, $\frac{z}{t}$ and all other ratios representing non homogeneous projective coordinates can be defined in the same way.

From this geometric definition, to reconstruct the projective coordinates of any matched pair of points is to be able to perform the basic geometric operation of intersecting a line with a plane only through image operations. This is possible once the epipolar geometry has been established.

Firstly, note that the viewing plane of the M_4M_5 with respect to the first image intersects the plane $M_1M_2M_3$ in the line L that goes through the point A. The projection l of L in the first image is confused with that of M_4M_5 in the first image plane but not in the second. In addition its projection in the second image can be reconstructed. If we note that N_4 (resp. N_5) is the intersection point of the viewing line of the M_4 (resp. M_5) with respect to the first image and the plane $M_1M_2M_3$. The line L is determined by N_4 and N_5. n_4 and n_5 are confused with m_4 and m_5 in the first image, and n'_4 and n'_5 can be reconstructed in the second image with the help of the epipolar geometry. Thus $\mathbf{l}' = \mathbf{n}'_4 \times \mathbf{m}'_5$ and the projection

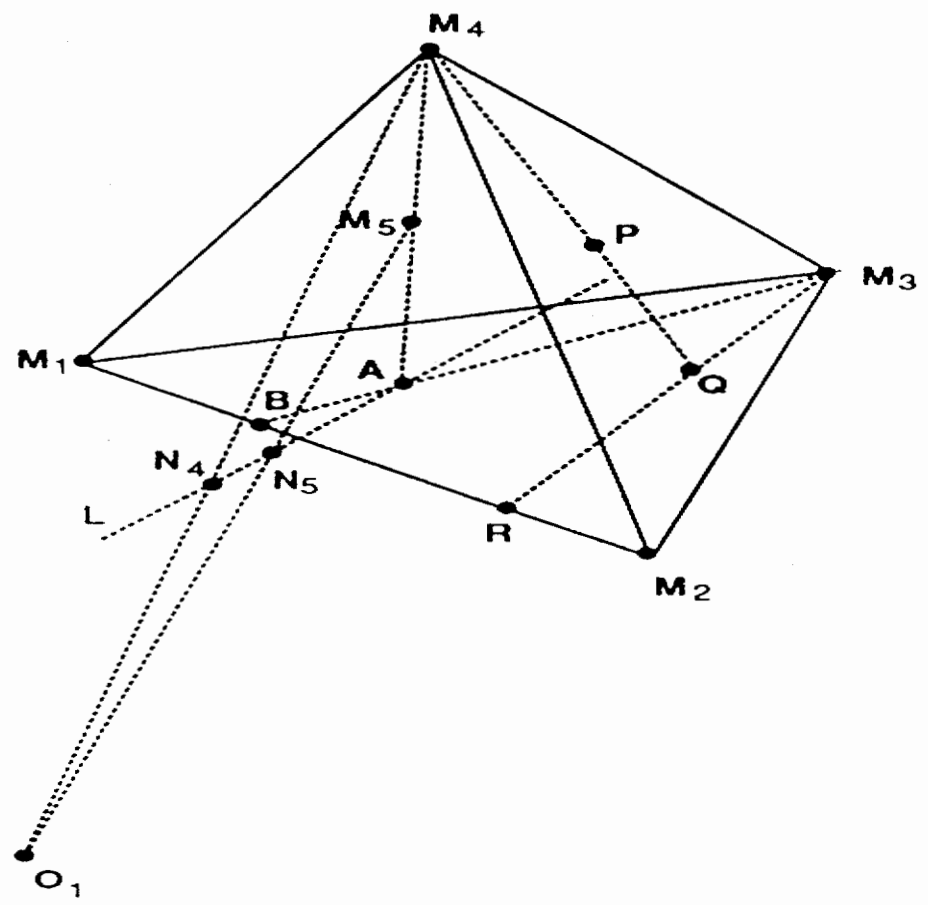

Figure 5: The geometric definition of projective coordinates.

a' of A in the second image is the intersection of l' and l'_{45} with l'_{45} is the line going through m'_4 and m'_5: $\mathbf{l}'_{45} = \mathbf{m}'_4 \times \mathbf{n}'_5$.

Therefore the projection b' of B in the second image is the intersection point of the line $a'm'_3$ and the line $m'_1m'_2$. In the same way as for M_5, we can deal with P to get the projection r' of R in the second image, the cross ratio of m'_1, m'_2, b', r' defines one of the non homogeneous projective coordinates of P, since we always have

$$\{m'_1, m'_2; b', r'\} \equiv \{M_1, M_2; B, R\}.$$

4 Practical reconstruction

For direct solution of the system 4 by non linear optimisation, as we have mentioned above, Levenberg-Marquardt's method is used. Practical experimentation shows that the algorithm works very well. The convergence does not depend too much on the initial starting points, it converges with almost *any initialisation* although we have no mean to formally prove its convergence.

To have some intuitive ideas of different shape representation, we will first show some reconstruction on a simulated cubic grid and a simulated glass (from the Gnuplot demonstration data) data. A projective shape is defined up to a collineation, no metric information is present, only projective properties are preserved. For example, aligned points remain aligned, coplanar points remain coplanar and conics are transformed into conics, a circle may be represented by an hyperpole ... Such a pure projective shape can be displayed by its non homogeneous coordinates (cf. Figure 6, Figure 7, and Figure 8).

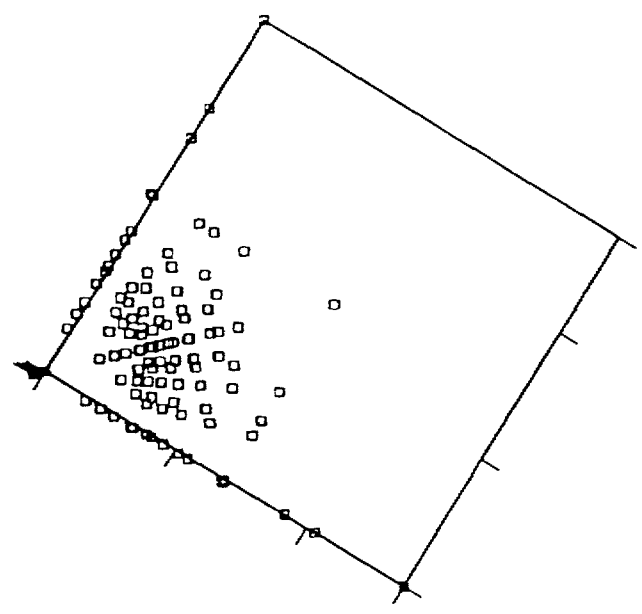

Figure 6: A *top* view of the projective "cubic grid".

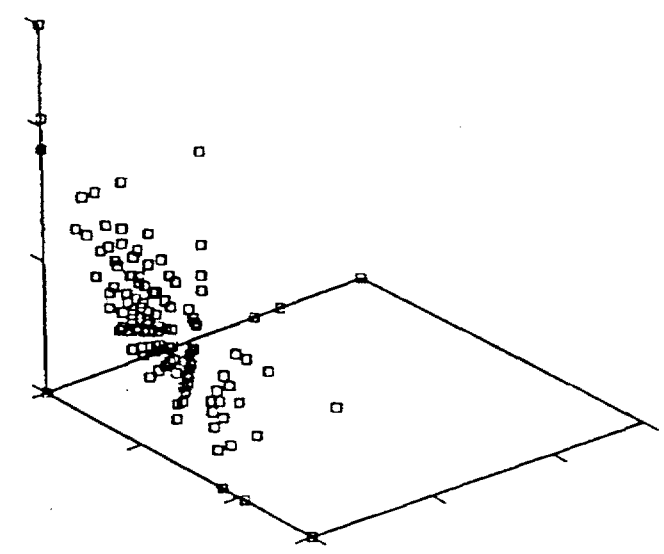

Figure 7: Another view of the projective "cubic grid".

Next, a pure projective shape can be transformed into its affine or Euclidean representation. However to do this, supplementary affine and Euclidean information shoud be incorporated. That is, we should determine a collineation $\mathbf{A}$ which brings the canonical basis $\mathbf{e}_i, i = 1, \ldots, 5$ to any five points

$$\mathbf{a}_i = (a_{i1}, a_{i2}, a_{i3}, a_{i4})^T = \mathbf{A}\mathbf{e}_i \tag{7}$$

If these five points are affinely known, that is 4 of them can be assigned the standard affine coordinates, the fifth point should have its affine coordinates with respect to these 4 points, that is the 5 points can have the

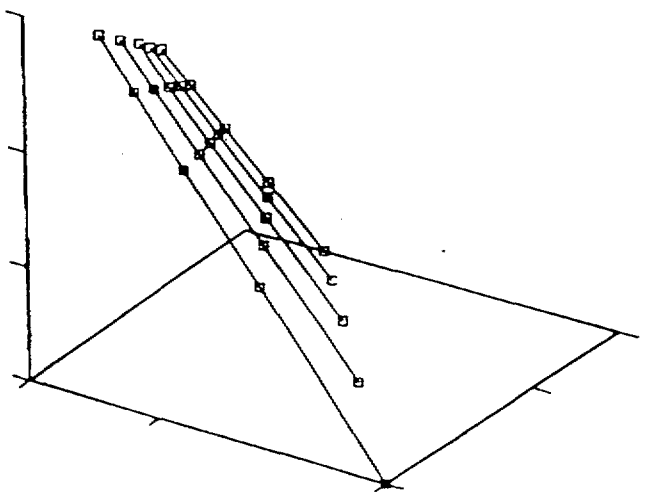

Figure 8: A projective plane.

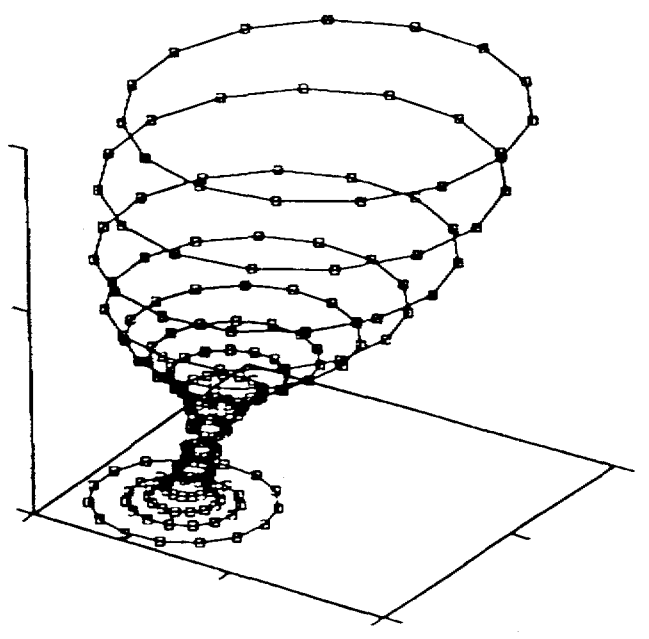

Figure 9: An affine reconstruction of a glass

following coordinates

$$(1,0,0,1)^T, (0,1,0,1)^T, (0,0,1,1)^T, (1,1,1,1) \text{ and } (\alpha,\beta,\gamma,1)^T$$

That is, to get the affine representation, affine knowledge (α,β,γ) have to be available. Then by solving the linear equations system 7, we obtain the collineation which transforms a pure projective shape into an affine shape.

To have the usual Euclidean shape representation, the Euclidean coordinates should be known for the 5 points, that should be like

$$(x_i, y_i, z_i, 1)^T, \quad i = 1, \ldots 5$$

then, solve for the corresponding collineation which transforms a pure projective shape into an usual Euclidean shape.

Figure 9 shows an affine shape of the glass.

We have also experimented on real data. All our experiences are conducted with a Pulnix 765 camera, a lens of $18mm$ and FG150 Imaging technology grab board. The camera is assumed to be a perfect pin-hole one, distorsion was not compensated. The first data set is obtained from the standard calibration pattern. Since the geometry of the pattern is regular, it makes possible the matching of the points, and the 3D measures are

available to verify the reconstruction results. Since only one planar pattern is available, we create a *"transparent"* pattern, that is, once the camera is fixed in a position, the pattern plane is then translated. This is equivalent to have several transparent calibration pattern. In our experimentation, we used 3 transparent planes and 3 positions (cf. Figure 10 and Figure 11). The contour points are obtained by a standard gradient based edge detector. Then follows the edge linking to obtain the least squares fitted lines. The image points are computed as the intersection points of the lines.

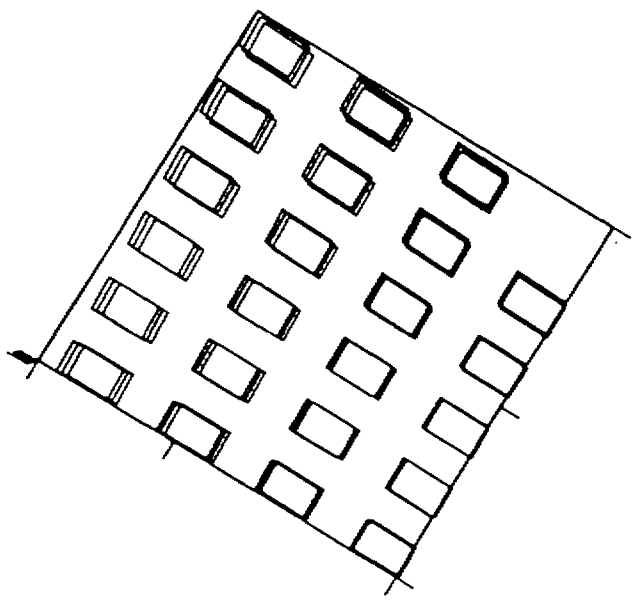

Figure 10: A top view of the reconstructed transparent calibration pattern

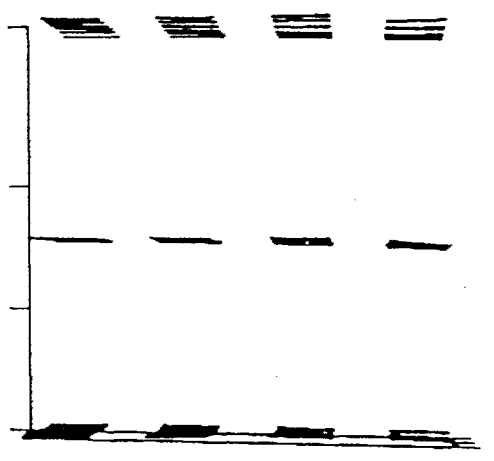

Figure 11: A side view of the reconstructed transparent calibration pattern

Another data set is obtained from a paper house. About 90 images are taken around the house. Then the curvature maxima are tracked by correlation operator. About 40 points are tracked over the total sequence. Then, five images of the total sequence are selected to perform the reconstruction.

In Figure 12 and 13, the first and the last images of the sequence are displayed.

Figures 14, 15 and 16 show the reconstructed house by the direct non linear method. Notice in these figures the quality of the reconstruction: windows are almost perfectly aligned with the wall. The boundary of the windows look like not lined up each other, they are really not in practice!

To estimate the confidence limits of the reconstructed points, we get the

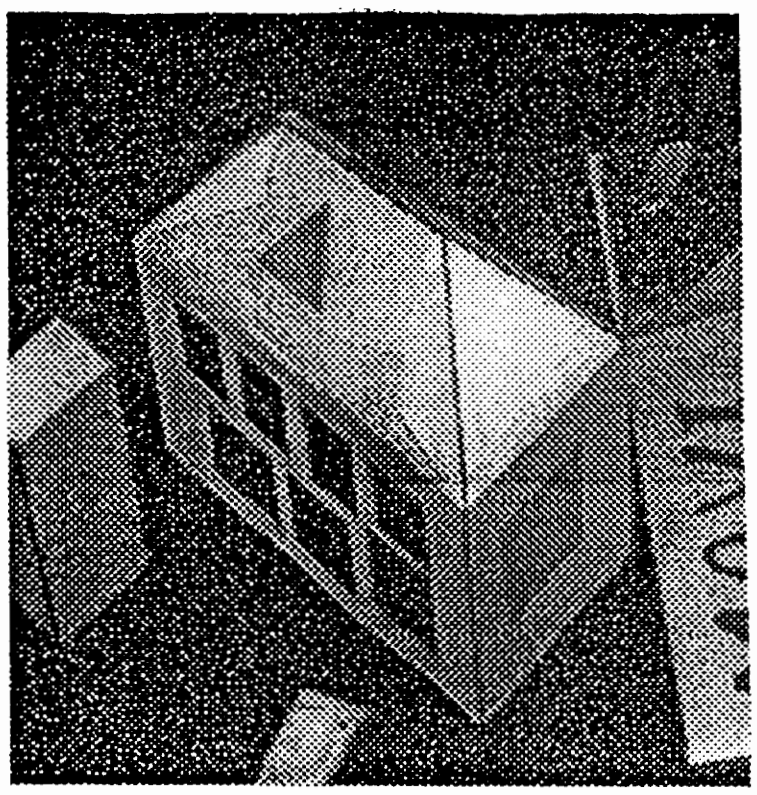

Figure 12: The first image of the sequence

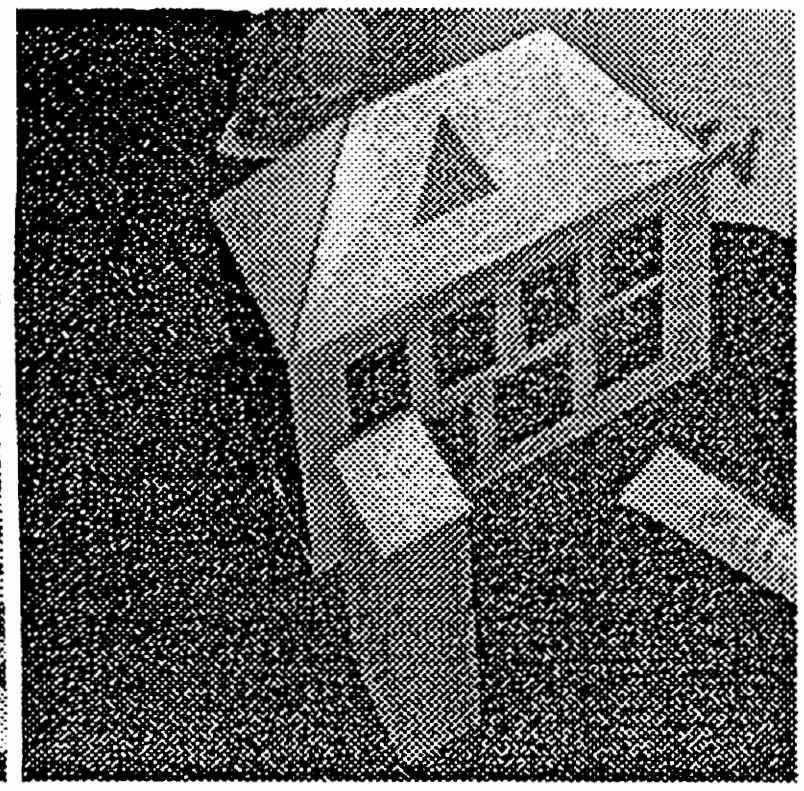

Figure 13: The last image of the sequence

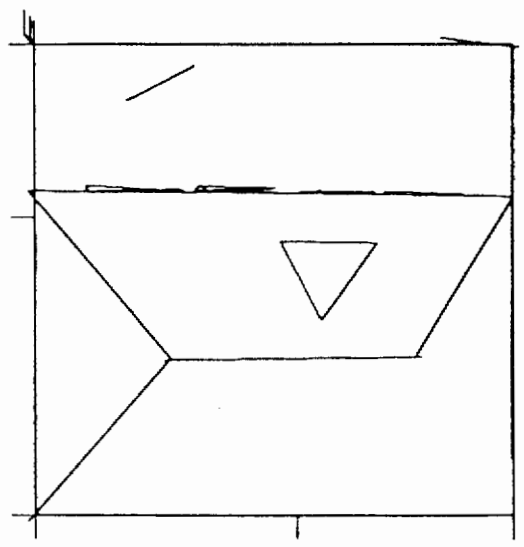

Figure 14: A top view of the reconstructed house by the direct method

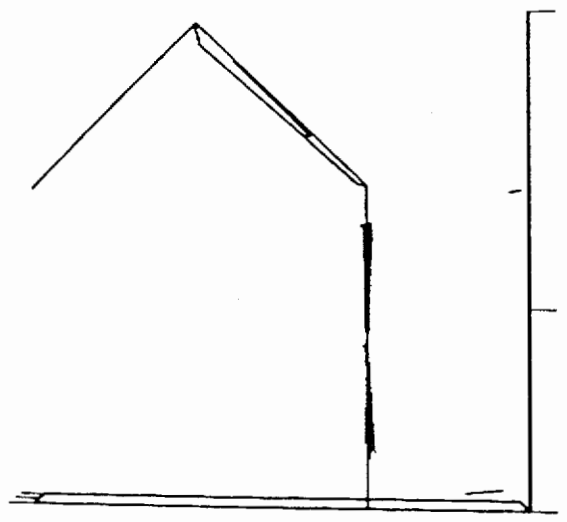

Figure 15: A side view of the reconstructed house by the direct method

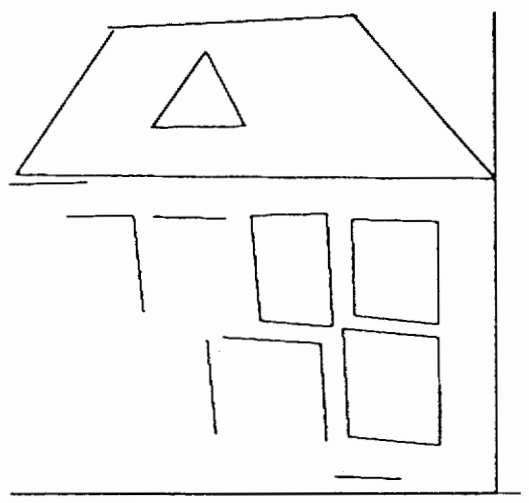

Figure 16: A front view of the reconstructed house by the direct method

covariance matrix from Levenberg-Marquardt's algorithm and we draw the confidence region ellipsoid related to each point as illustrated by the figures 17, 18 in which each associated ellipsoid is displayed by its corresponding bounding parallelepiped.

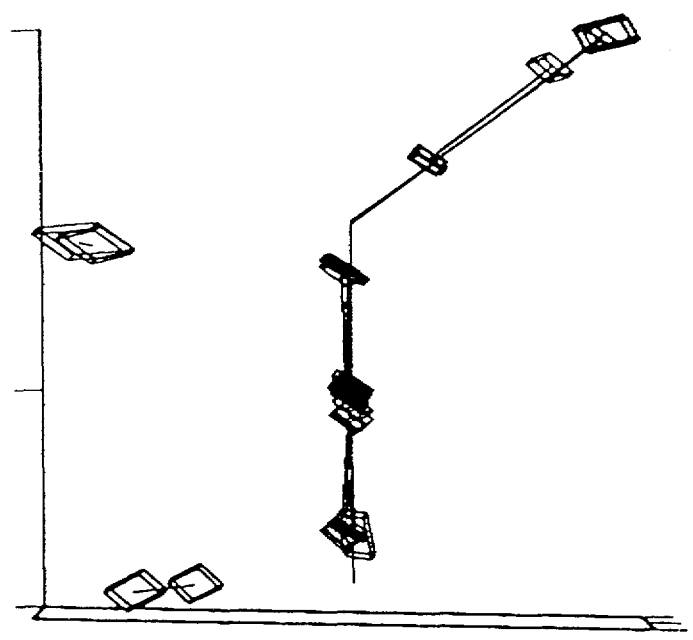

Figure 17: A side view of the reconstruction with their confidence region.

5 Discussion

As concluding remarks, the relative reconstruction is of high quality, having very good numerical behavior, this kind of relative approach has many advantages over the traditional approaches both mathematically and experimentally. One of the most important key points of the approach is that we have no need to calibrate cameras. Therefore zooming, focusing ... of an active moving camera can be fully absorbed by this calibration free process. Multiple observation mesures are natually globally integrated in the estima-

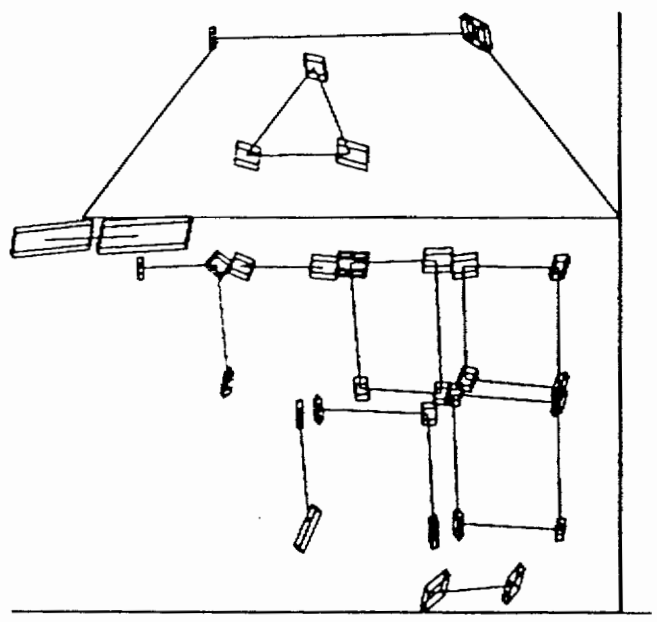

Figure 18: A front view of the reconstruction with their confidence region

tion process, though scene points should not necessaryly be present in all images, thus provides us with a more precise reconstruction. Moreover the mathematical formulation appears in an extremely simple form thanks to the elegant projective geometry.

As for non linear least squares estimation, Levenberg-Marquardt's algorithm provides actually quite satisfactory results. The convergence does not depend on the initial starting point and the number of iterations necessary is reasonable. However specialied implementation of Levenberg-Marquardt for our problems will improve the computing time, and also some more powerful numerical optimisation algorithm such as confidence region can be used to still improve the results.

We are currently beginning on the studies of the precision of the reconstruction with a full statistical model following our first experiments illustrated in Figures 17 and 18. We are also comparing the method with the others dealing with the similar problems, especially the linear method of Faugeras [5], but it is too early for the time being to make conclusion on these comparative studies.

Acknowledgement

We are pleased to acknowledge Horst Beyer for pointing out to us that photogrammetrists use Levenberg-Marquardt for closely related problems, Olivier D. Faugeras and his associates for fruitful discussions on the topic and Radu Horaud for helping us to prepare calibration pattern data.

References

[1] H.A. Beyer. Accurate calibration of CCD cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, page to appear, 1992.

[2] D.C. Brown. Close-range camera calibration. *Phtogrammetric Engineering*, 37(8):855–866, 1971.

[3] J.B. Burns, R. Weiss, and E.M. Riseman. View variation of point set and line segment features. In *Proceedings of DARPA Image Understanding Workshop, Pittsburgh, Pennsylvania, USA*, pages 650–659, 1990.

[4] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[5] O.D. Faugeras. *3D Computer Vision.* M.I.T. Press, 1992.

[6] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera Self-Calibration: Theory and Experiments. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 321–334. Springer-Verlag, May 1992.

[7] O.D. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. In *IEEE workshop on Computer Vision*, 1989.

[8] P. Gros and L. Quan. Projective Invariants for Vision. Technical report, IRIMAG–LIFIA, Grenoble, France, 1992.

[9] R.I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 579–587. Springer-Verlag, 1992.

[10] J.J. Koenderink and A. J. van Doorn. Affine structure from motion. Technical report, Utrecht University, Utrecht, The Netherlands, October 1989.

[11] R. Mohr and L. Morin. Relative positioning from geometric invariants. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, USA*, pages 139–144, June 1991.

[12] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling W.T. *Numerical recipes in C.* Cambridge University Press, 1988.

[13] L. Quan and R. Mohr. Affine shape representation from motion through reference points. *Journal of Mathematical Imaging and Vision*, 1:145–151, 1992.

[14] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry.* Oxford Science Publication, 1952.

[15] G. Sparr. Projective invariants for affine shapes of point configurations. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Reykjavik, Iceland*, pages 151–170, March 1991.

[16] G. Sparr and L. Nielsen. Shape and mutual cross-ratios with applications to the interior, exterior and relative orientation. In *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pages 607–609. Springer-Verlag, April 1990.

[17] C. Tomasi and T. Kanade. Factoring image sequences into shape and motion. In *Proceedings of IEEE Workshop on Visual Motion, Princeton, New Jersey*, pages 21–28, Los Alamitos, California, October 1991. IEEE Computer Society Press.