Relative 3D Reconstruction Using Multiple Uncalibrated Images

Roger Mohr, Long Quan, Françoise Veillon

Relative 3D Reconstruction Using Multiple Uncalibrated Images

R. Mohr L. Quan F. Veillon

LIFIA - CNRS - INRIA,
46, avenue Félix Viallet,
38031 Grenoble, France

International Journal of Robotics Research, Vol. 14, No. 6, PP. 619–632, 1995

Abstract

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

Key words: *relative reconstruction, projective geometry, uncalibration, geometric interpretation*

1 Introduction

1.1 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 as shown by Burns, Weiss and Riseman (1990). 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 absolute reference frame.

This approach suffers from several drawbacks: firstly the calibration process is an error sensitive process; secondly it cannot always be performed on 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 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: photogrammetrists as Brown (1971) and Beyer (1992) use to solve calibration and reconstruction in the same process. 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, for instance in Mohr and Arbogast (1990), Mohr et al. (1991), Lee and Huang (1990), Tomasi and Kanade (1991), Kœnderink and van Doorn (1989), Sparr (1991).

This year, three independent teams approached the same problem of 3D reconstruction from uncalibrated cameras, and all three with the same projective basis. Faugeras (1992) published an insightful algebraic method to perform 3D projective reconstruction with the tricky use of the epipolar geometry of an image pair. He demonstrated that once the epipolar geometry is somehow determined, 3D projective structure can be reconstructed up to a collineation by assigning 5 reference points to the standard projective basis. One month later, Hartley et al. published their paper (1992) which is a shortened version of an extended report available at G.E.C. In this paper, he describes a similar approach in a slightly different way. At the same time appeared our technical report (Mohr et al. 1992) where our first experimental results are presented.

The key idea of this kind of approach is to take the reference points of a scene neither in a camera reference frame nor an absolute reference frame, but points whose 3D positions are unknown, but which are located in the images. First, exploratory work was done in the case of true perspective projection (Mohr 1990, Mohr 1991). Much more interesting was the approach taken by Kœnderink and Van Dorn (1989) and independently by Lee and Huang (1990) and Tomasi and kanade (1991) for the affine case of orthographic projection. In this case, only four reference points to define an affine frame are necessary instead of 5 for general perspective projection. Also the basic recovered 3D structure is an affine shape in this case instead of a projective shape.

The original contributions of this paper are mainly twofold. First, we describe a direct 3D relative reconstruction method, which differs from that of Faugeras (1992) and Hartley et al. (1992) in that our method is formulated globally as a least squares estimation method which does not need to first estimate the epipolar geometry; also the method makes full use of redundancy of multiple images. However, both Faugeras and Hartley et al.'s methods depend essentially on the success of the determination of the epipolar geometry from the fundamental matrix. The work described in this paper is related to recent work in the photogrammetry community on self calibration method (Beyer 1992). Both calibration and reconstruction are incorporated in the same optimization process. We describe an implementation which allows to solve the problem in presence of noise, using redundant data. This is done by an implementation of parameters estimation theory, using Levenberg-Marquardt algorithm. We will also discuss the robustness of the method, and the precision of reconstruction from experimental results on both real and simulated image sequences.

Secondly, we provide a geometrical way to choose among the set of points those which can be selected as reference points. The selected reference points should not be degenerated, i.e. no four of them coplanar. This result allows first to derive a computational way to choose the correct reference points and secondly to provide a geometrical interpretation of relative reconstruction (Mohr et al. 1992) as in Koenderink and Van Doorn (1989) for affine case and

our previous work in Quan and Mohr (1991).

1.2 Context of FIRST project

One of FIRST primary goals was to improve robustness of sensing within the context of robotic planning. In such a context we developed an object-based way to deal with 3D perception from multiple images. Classical sensing methods heavily rely on off-line calibration both for cameras and for the hand-eye link. Such a calibration is often difficult and noise sensitive as described in the previous section.

Telative positioning is an interesting alternative for solving this problem of 3D perception as it bypasses calibration, using invariant properties of the imaging process. In our case, object reconstruction is relative to reference points selected in the images and for which the 3D quantitative nature can remain totally unknown. The structure of the objects reconstructed in such a way is first an invariant projective representation. From there, some direct relative information can be already derived. For instance any point is lying *on* or *above* or *below* a three points defined plane. All this needs no 3D quantitative information. Later on, different levels of 3D knowledge may be incorporated, leading to affine or Euclidean reconstruction.

This approach has two major advantages in this robot planning context: Firstly it introduces the uncertain 3D information only at the latest stage of the perception process, and only if it is needed. Secondly it avoids the unstable calibration process and therefore errors relies only on the accuracy of the measures in the image during the processing.

In the same project, the Oxford team developed an object recognition system based on a similar approach using projective invariants (Forsyth et al. 1991).

1.3 Outline of the paper

The paper is organized as follows. First, section 2 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 we show how 3D reconstruction is reached from multiple uncalibrated images. Section 3 provides basic results on the automatic checking of coplanar points through epipolar geometry. Then we describe how reference points can be automatically selected. In section 4, experimental results will be presented and discussed. Section 5 concludes the paper by a discussion on the subject and some future works.

Two basic assumptions are made throughout the paper. First we assume that the reader is familiar with elementary projective geometry, as it can be found in the first chapters of Semple and Kneebone (1952) and also Faugeras (1993). We also assume 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 reference points

This section provides the basic equations of the 3D reconstruction problem, together with that of self calibration. This derivation was developed independently from those recently published

by Faugeras (1992). The basic starting point is similar to this work; however the way to solve it was influenced by the way photogrammetrists simultaneously calibrate their camera and reconstruct the scene, by use of carefully located beacons (Beyer 1992).

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 only 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 $(u_{ij}w_{ij}, v_{ij}w_{ij}, w_{ij})^T$ or its usual non homogeneous coordinates $(u_{ij}, v_{ij})^T$. 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 which is 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:

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

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

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 redundant set of equations.

It is easy to understand that the solution of equation 1 is not unique. For instance, if the origin is translated, all 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 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 3D space. For a projective space $I\!P^3$, 5 algebraically free points form a basis, i.e. 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 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 building an explicit solution from these now fixed reference points.

2.2 Direct nonlinear reconstruction method

From the above section, the most direct way is to try to solve this system of nonlinear 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 a standard least squares technique. The problem can be formulated as minimizing

$$F(x_i, y_i, z_i, t_i, p_{11}^{(j)}, ...p_{34}^{(j)}) = \sum_{k=1}^{2\times m\times n+n} (\frac{f_k(u_{ij}, v_{ij}; x_i, y_i, z_i, t_i, p_{11}^{(j)}, ...p_{34}^{(j)})}{\sigma_k})^2$$

over

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

where $f_k(\cdot)$ is either

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

or

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

subject to

$$x_i^2 + y_i^2 + z_i^2 + t_i^2 - 1 = 0 \text{ for } i = 1, \ldots, m.$$

σ_k is the standard deviation of each image measure, u_{ij} or v_{ij}, suppposed normally distributed and uncorrelated. On the other hand, it can also be considered as a weight for each function. So the problem is a general weighted least squares estimation; thus the constraints $x_i^2 + y_i^2 + z_i^2 + t_i^2 - 1 = 0$ can be easily transformed into corresponding penalty functions in order that the whole problem is an unconstrained least squares problem.

As for the multiplicative scalar of each projection matrix, we can for example impose $p_{34}^{(j)} = 1$ for $j = 1, \ldots, n$ for general real camera positions.

The only known measures are the image points (u_{ij}, v_{ij}). All others are unknown parameters to estimate. This system leads to $n + 2 \times n \times m$ equations in $11 \times m + 3 \times n$ unknowns,

This can be solved by the standard nonlinear least squares routine due to Levenberg-Marquardt (Press, Flannery et al. 1988). Statistically, it is equivalent to the maximum likelihood estimator. Let $J(\cdot)$ be the Jacobian matrix of $F(\cdot)$. At each iteration k, the search direction $\delta \in I\!R^{11\times m+3\times n}$ is obtained by solving

$$(J_k^T J_k + \lambda_k)\delta = -J_k^T F_k$$

where λ_k is a non-negative scalar which will increase or decrease by a factor of 10 according to the increase or decrease of $F(\cdot)$. Thus the method is based on the quadratic modeling of the objective function. The Hessian matrix of $F(\cdot)$ is rather approximated by $J^T J$ than explicitly calculated. This method has some strong global convergence properties in practice.

The alternative of minimizing $F(\cdot)$ as above is to minimize

$$G(x_i, y_i, z_i, t_i, p_{11}^{(j)}, ...p_{34}^{(j)}) = \sum_{k=1}^{2\times m\times n+n} (\frac{g_k(u_{ij}, v_{ij}; x_i, y_i, z_i, t_i, p_{11}^{(j)}, ...p_{34}^{(j)})}{\sigma_k})^2$$

where $g_k(\cdot)$ is either

$$u_{ij}(p_{31}^{(j)}x_i + p_{32}^{(j)}y_i + p_{33}^{(j)}z_i + p_{34}^{(j)}t_i) - (p_{11}^{(j)}x_i + p_{12}^{(j)}y_i + p_{13}^{(j)}z_i + p_{14}^{(j)}t_i)$$

or

$$v_{ij}(p_{31}^{(j)}x_i + p_{32}^{(j)}y_i + p_{33}^{(j)}z_i + p_{34}^{(j)}t_i) - (p_{21}^{(j)}x_i + p_{22}^{(j)}y_i + p_{23}^{(j)}z_i + p_{24}^{(j)}t_i).$$

$g_k(\cdot)$ is a simple algebraic transformation of $f_k(\cdot)$. This transforms the real Euclidean *distance* error into an *algebraic distance* which degrades the error function. However, doing so, the degree of nonlinearity of equations is greatly reduced, especially the Jacobian matrix of $g_k(\cdot)$ is nicely reduced. This may lead to faster convergence but leaves the solution a little bit degraded, since the distance error is only algebraic, not Euclidean. This point will be discussed later and get confirmed in our experimentation in section 4.

Since the standard projective basis are assigned to the reference points, the solution provides at the same time the projective shape and each camera's projection matrix. A projective shape is defined up to a collineation. At this stage, 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 hyperbole, and so on.

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. We have to determine a collineation $\mathbf{A}$, a 4×4 matrix which brings the canonical basis $\mathbf{e}_i, i = 1, \ldots, 5$ to any five points

$$\mathbf{a}_i = (a_{i1}, a_{i2}, a_{i3}, a_{i4})^T = \mathbf{A}\mathbf{e}_i \tag{4}$$

If these five points are only affinely known, i.e. 4 of them can be assigned the standard affine coordinates, the coordinates of the fifth point must be affine coordinates with respect to these five points. Therefore the 5 points can have the following coordinates

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

Consequently, to get the affine representation, affine knowledge (α, β, γ) has to be available. Then by solving the linear equations system 4, we obtain the collineation which transforms a pure projective shape into an affine shape.

To obtain the usual Euclidean shape representation, the Euclidean coordinates have to be known for the 5 points, for instance:

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

Then, with these 5 points we can compute the corresponding collineation which transforms a pure projective shape into an usual Euclidean shape.

However we can also assign the reference points to their Euclidean coordinates at the beginning of the minimization process. In this case, the 3D reconstruction thus obtained is directly its Euclidean shape.

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 operations in the image planes. That leads to an automatic selection of general reference points from image planes and point reconstruction in a geometric way. For more details concerning the geometric interpretation of projective reconstruction from two images (Mohr et al. 1992).

The computation of the fundamental matrix is done by non linear optimization method as proposed in (Faugeras et al. 1992). It is important to note that for projective reconstruction, the fundamental matrix is not necessary at all; it is only used for selecting correct reference points.

3.1 The coplanarity test

As we assume here that the epipolar constraint is known, we know the fundamental matrix F which contains all this information (Faugeras 1992; Hartley et al. 1992). $\mathbf{F}$ 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 the image 2 has its coefficients satisfying $\mathbf{l'} = (a', b', c')^T = \mathbf{Fm}$.

Now, consider Figure 1. It displays the projections of four 3D points A, B, C, D in 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. 1.

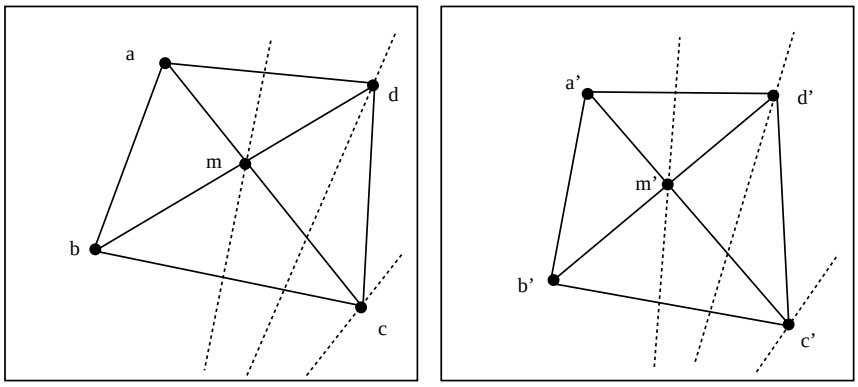

Figure 1: Match of diagonal intersections with epipolar constraint

Conversely consider the case where A, B, C, D are not coplanar. Diagonals are no more in the same plane and therefore do 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 plane (BDO); then the 2 view lines (Om) and $(O'm')$ do not intersect, and therefore the points m and m' are not in epipolar correspondence.

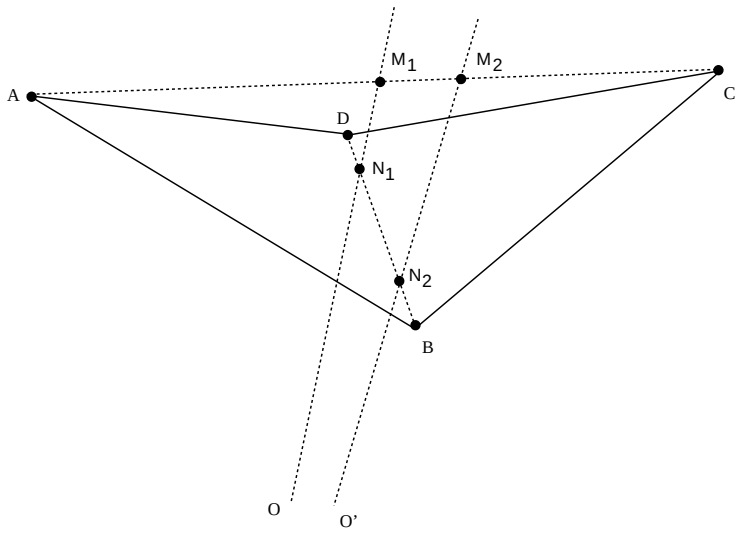

Figure 2: Four non coplanar points in space

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 easily checked. 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 neither $a, b, c,$ 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.*

Thus it allows us to check if any given 4 points are coplanar.

3.2 Search for a 5 point basis

The above result can directly be used to automatically select, from image points, the reference points necessary for projective reconstruction, without any *a priori* spatial knowledge. Basically, the previous section results allow us to get rid of the coplanar reference points in the step 3 of the following algorithm.

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 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 other points according to *distances* to the plane determined by triangle $M_1M_2M_3$, choose for M_4 the one which has the maximum distance. The *distance* is not the real orthogonal distance from the point to the plane (not calculable at this step), it is the projection on the second image of the segment from the point to the plane along the first viewing line of that point (see Figure 3),
4. Sort remaining points according to the maximum distance to any face of tetrahedron M_1, M_2, M_3, M_4, choose for M_5 the point which has the maximum distance.

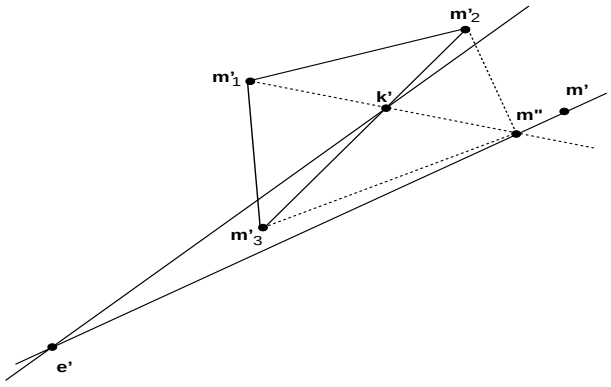

Figure 3: The *distance* is defined as that between m' and m''.

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

4 Experimental results

4.1 Qualitative results

All our experiences are conducted with a Pulnix 765 camera, a lens of $18mm$ kinoptics and FG150 Imaging technology grab board. The camera is assumed to be a perfect pin-hole one, distorsion is not compensated. The first data set has been obtained from a paper house of about 30 centimeters large which was placed at about 1.50m from the camera. Then nearly 45 images were taken around the house covering roughly 90 degrees of rotation angle. Each successive pair of images are close enough to facilitate tracking of points of interest. Contour points of each image are obtained by a Canny-like edge detector. Then follows the linking of contour points. Each contour chain is approximated by a regularized cubic B-Spline curve. The curvature maxima of a B-Spline curve are considered as points of interest. They therefore are tracked over the total image sequence. Tracking is based on the correlation of points of interest between successive images. About 40 points are in this way tracked over the total sequence. Then, only five images of the total sequence are selected to perform reconstruction.

Figure 4 shows the first and the last image of the sequence.

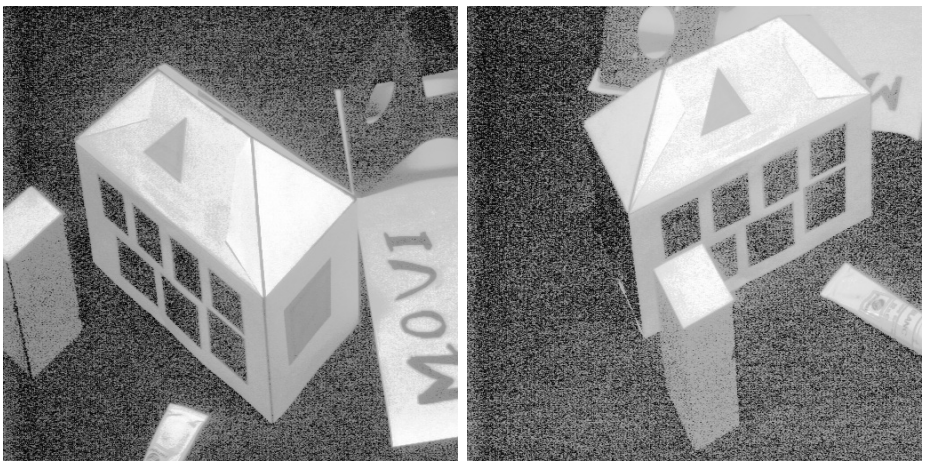

Figure 4: The paper house image sequence

Figure 5 shows the reconstructed house. Notice in these figures the quality of the reconstruction: windows are almost perfectly aligned with the wall. The boundary of the windows looks like not lined up each other, they are really not in practice!

Please note that after projective reconstruction, the projective transformation which transforms the projective reference points into their known Euclidean coordinates is applied to the projective shape in order to be displayed.

Another experience is performed on a wooden house. The wooden house is a little bigger than the paper house, the camera is set at about 2m from the object. We tracked over about one hundred images covering roughly two sides of it. In this experience, we wanted to validate the reconstruction with points only present in part of the sequence. In total, 73 points were tracked, but a small half of them are present between two successive views. Final reconstruction is done with five views of them.

In Figure 6, three images of the sequence are displayed.

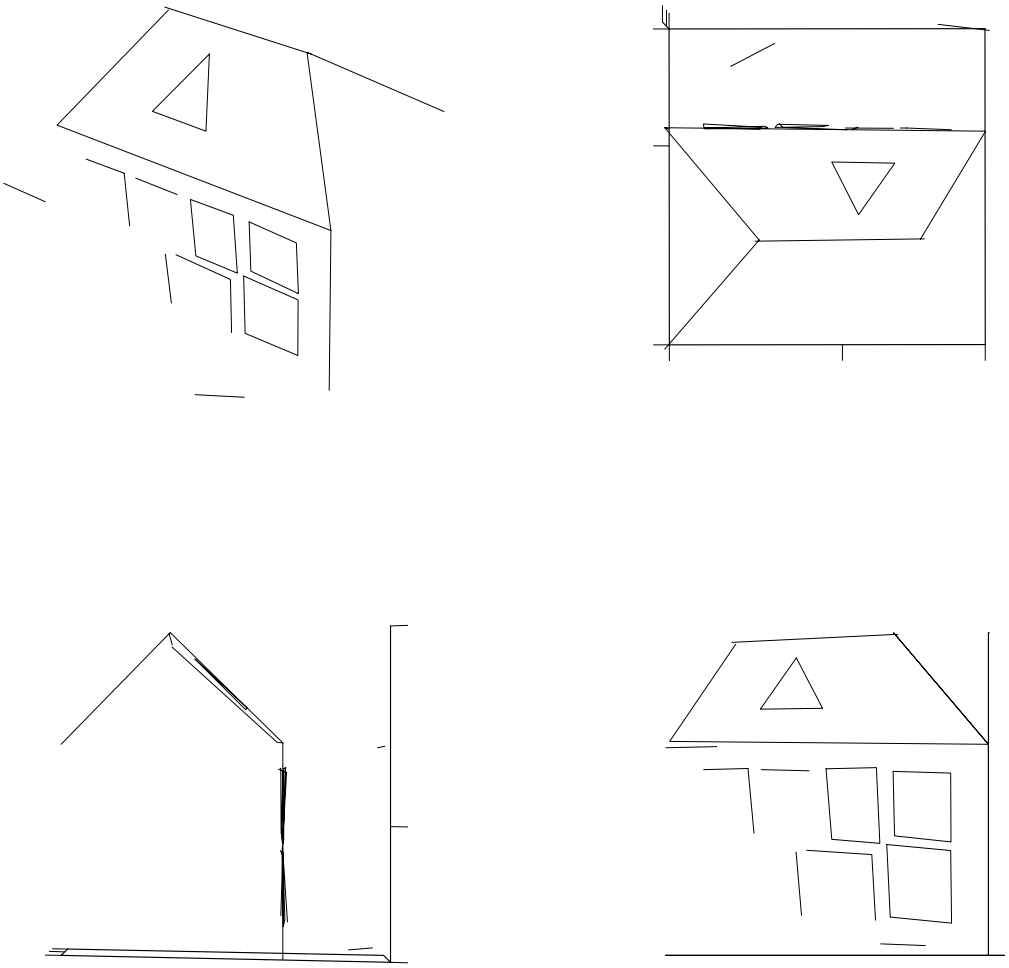

Figure 5: Some selected views of the reconstructed paper house

Figure 6: The wooden house image sequence

The reconstruction, illustrated in Figure 7, has an excellent qualitative aspect.

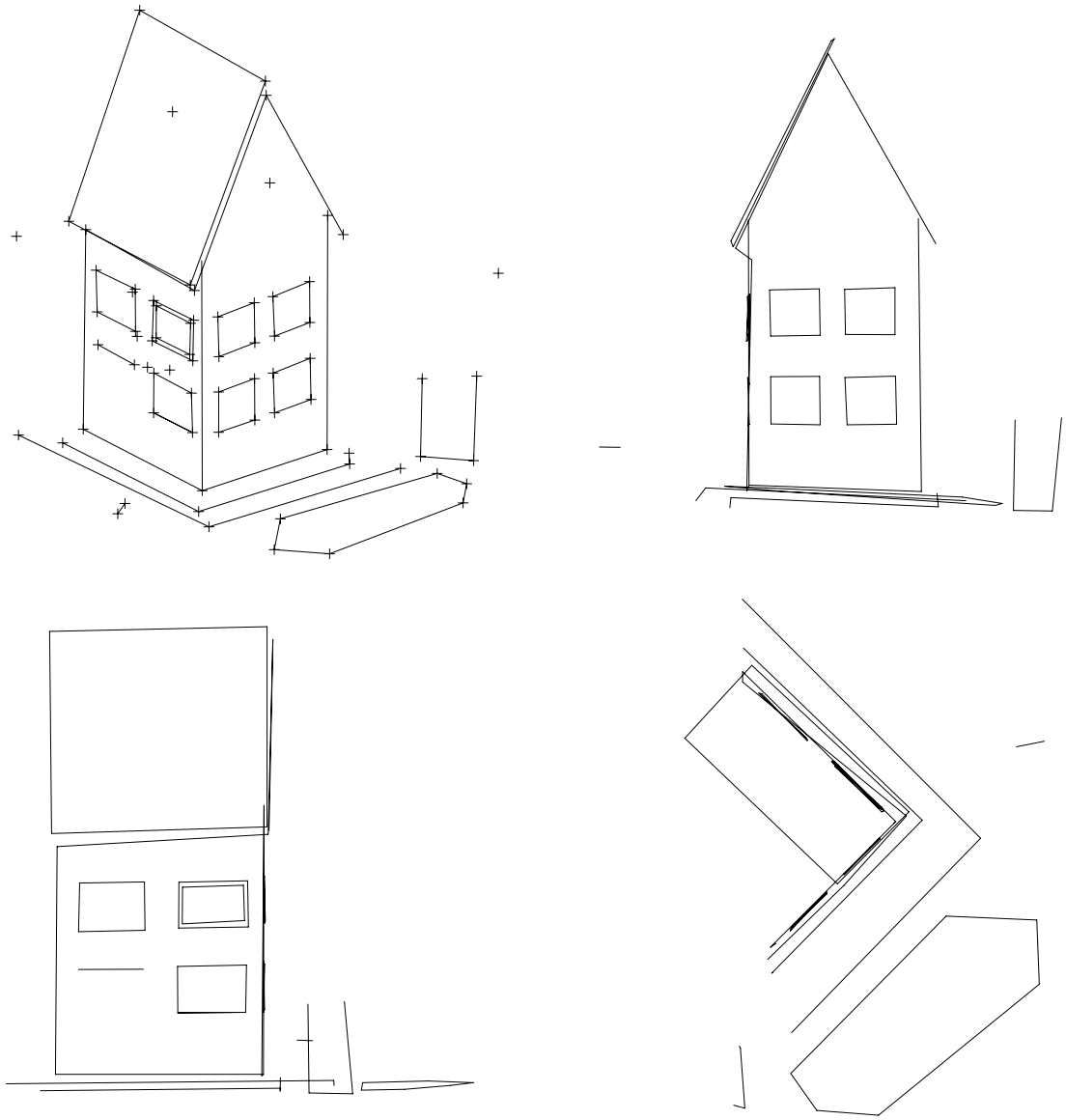

Figure 7: The reconstructed wooden house

As previously mentionned, we have choice for either minimizing $F(\cdot)$ or $G(\cdot)$. Experiences confirm that while minimizing $G(\cdot)$ with very few iterations (about 5 instead of 10 or more), we can obtain a quite satisfactory solution. But since the distance error is only algebraic, not Euclidean, the solution is always slightly degraded. In our experiments, we began with minimizing $G(\cdot)$, and ended with minimizing $F(\cdot)$.

All experimental results are performed by Levenberg-Marquardt's algorithm. Practical experimentation shows that the algorithm works very well. The convergence does not depend too much on the initial starting points. From our experiments, it came out that the initial data for the 5 reference points should be the coordinates

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

and that they roughly correspond to the configuration a) of Fig. 8 of similar position in the space, but that a strong wrong relative position of a point like M_5 in the case b might make the system to diverge.

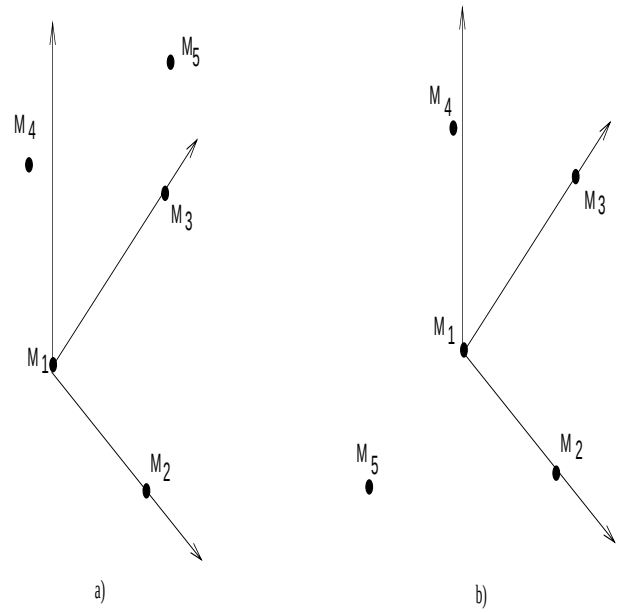

Figure 8: The configuration of the reference points.

Initialization of the projection matrices and points other than reference points proved to be little sensitive; a key point was to put enough high value in the first elements of the last column (see for a real camera where this componant comes from). For instance all our examples run with:

$$\begin{pmatrix} 0.1 & 0.1 & 0.1 & 20 \\ 0.1 & 0.1 & 0.1 & 20 \\ 0.1 & 0.1 & 0.1 & 1 \end{pmatrix}$$

and for points other than reference points:

$$(0.5, 0.5, 0.5, 1).$$

Unfortunately neither mathematical proof of convergence nor warranties for convergence can be provided. Practically convergence was obtained after five to ten iterations.

Some other experiments, for instance on the calibration patterns, are also performed and have been reported in the technical report. No convergence problem is encountered in our experiments.

4.2 Quantitative results

The accuracy of the tracked points is generally within two pixels, but some of them may bear more than that. To get an idea of the precision of the reconstructed points, we measured some points' coordinates of the wooden house by a ruler. Figure 9 shows the superimposition of the estimated points (transformed by the Euclidean coordinates of the 5 reference points) and the measured points.

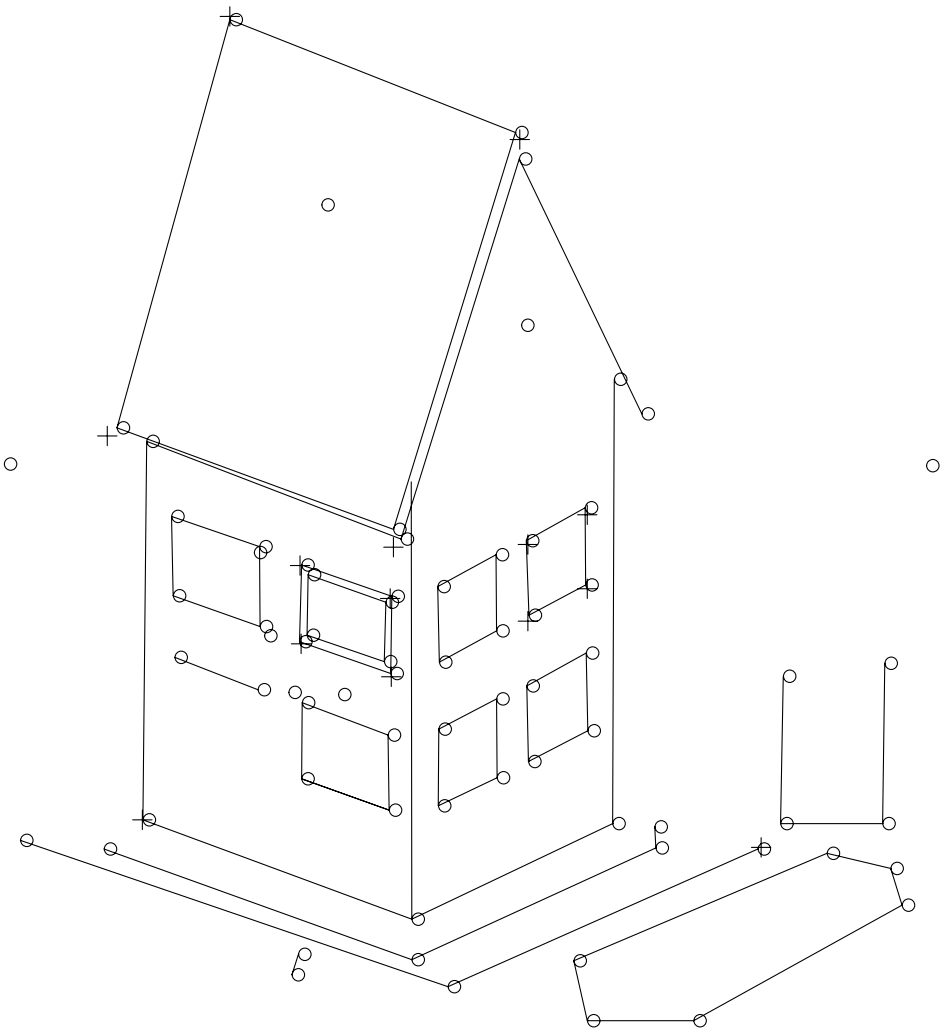

Figure 9: crosses represent some measured points, circles the reconstructed points

estimated coordinates	measured coordinates	absolute error
(11.678 -0.013 8.256)	(11.8 0 8.1)	(0.122 0.013 0.156)
(7.697 -0.035 10.663)	(7.85 0 10.5)	(0.153 0.035 0.163)
(6.646 -0.300 24.051)	(6.85 -0.4 23.8)	(0.204 0.1 0.251)
(11.666 0.007 10.766)	(11.8 0 10.5)	(0.134 0.007 0.266)
(7.773 -0.166 8.241)	(7.85 0 8)	(0.077 0.166 0.241)
(-0.065 1.300 7.922)	(0 1.35 7.8)	(0.065 0.05 0.122)
(-0.139 7.261 7.860)	(0 7.3 7.75)	(0.139 0.039 0.11)
(0.082 1.372 10.407)	(0 1.4 10.35)	(0.082 0.028 0.057)
(0.007 7.250 10.325)	(0 7.35 10.3)	(0.007 0.10 0.025)
(-1.488 -0.298 13.299)	(-1.7 -0.5 12.8)	(0.212 0.202 0.501)
(-1.086 18.143 12.934)	(-1.7 18.2 12.8)	(0.614 0.057 0.134)

Table 1: Absolute errors of the reconstructed points

The following numerical table 4.2 shows the absolute errors of the reconstruction of some selected points. While taking into account of rough measures' performance by the ruler, the absolute error is within one millimeter. It is a very acceptable result.

Because of lack of ground truth of a real object, simulated data were used to measure the precision of reconstruction. A uniformely distributed noise between $[-n, n]$ pixels is added to these simulated data. Reconstruction has been performed for different values of n.

Figure 10 shows the simulated data, superimposed with the reconstruction obtained with a noise such that $n = 2.5$. We have noticed that with one pixel noise, the difference of reconstruction is almost invisible.

Figure 11 illustrates the reconstruction precision according to the different pixels' noise.

As the least squares estimator can be considered as an maximum likelyhood one if we admit that the images points are normally distributed, that is what we assumed at the beginning. The confidence limits of the reconstructed points can be estimated from the corresponding covariance matrix provided by Levenberg-Marquardt's algorithm, the formula can be found in (Press et al. 1988). In Figure 12, the confidence region ellipsoid of each point corresponding to 68.3 percent confidence region is displayed. For simplicity, each associated ellipsoid is displayed by its corresponding bounding parallelepiped.

It is very important to note that in this figure the point with the largest confidence region is the point which lies on the plastic cup. Therefore it is not a real 3D "corner" in the original image. The true physical "corner" points have very small confidence regions.

5 Discussion

This paper presents a reconstruction method which can be easily implemented. One of its most important feature is that it does not require camera calibration. Therefore zooming, focusing $\cdots$ etc of an active camera are naturally integrated in the estimation process. A projective basis of five reference points are used in this method and reconstruction is performed within this frame. A geometric method is provided for selecting these reference points from only two images.

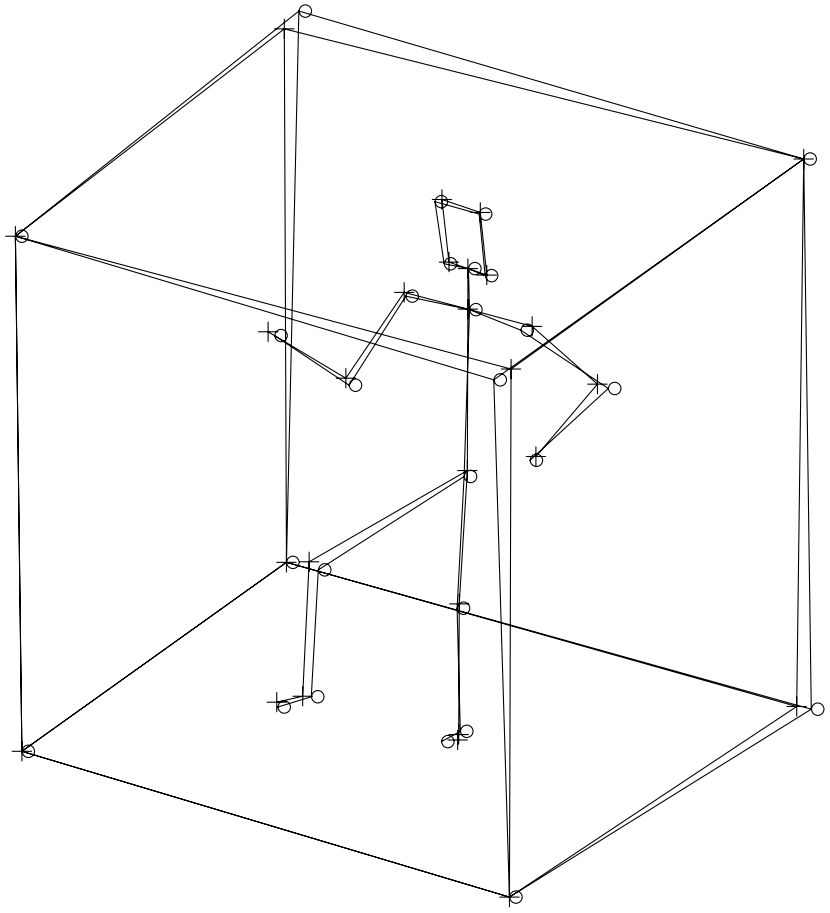

Figure 10: The reconstruction of simulated data: crosses represent the initial simulated points, circles the reconstructed points from 5 pixels noise

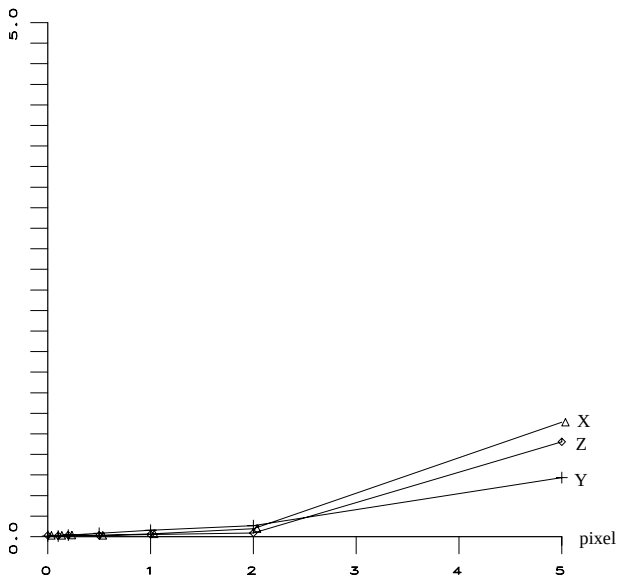

Figure 11: Relative errors (in percentage) variation according to pixel perturbation

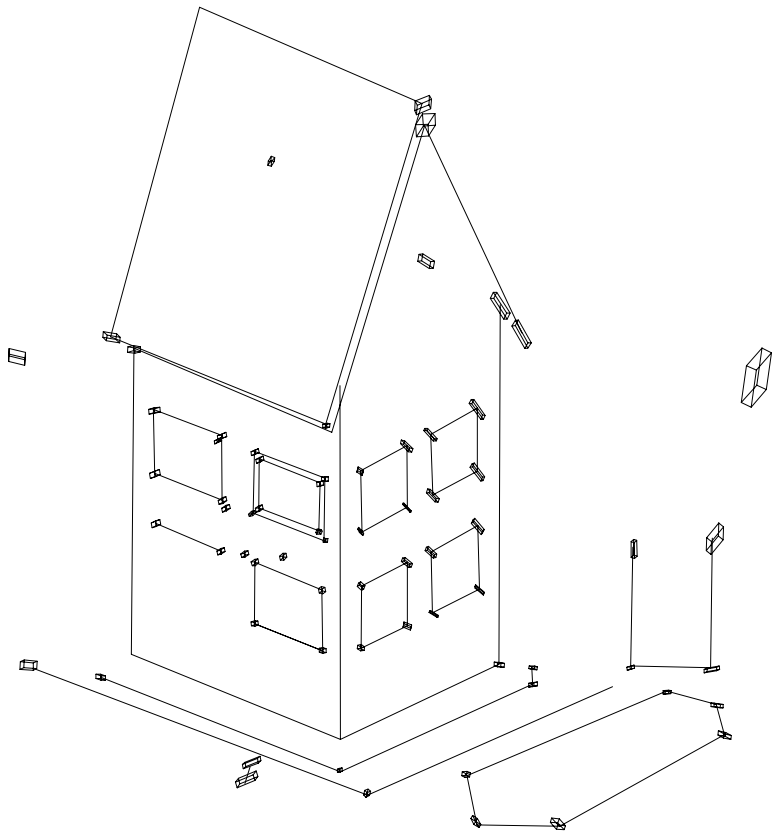

Figure 12: Confidence regions

The qualitative results are excellent. If we assume that the exact location of the reference points are known, quantitative results can be obtained; they are better than those provided by stereovision, but still not excellent enough for accurate industrial applications. One first way to improve the location accuracy is to have better location measures in the images. Presently we locate corners in a simple way: from a Canny like contour extractor, B-splines are fitted using least-squares approximation. Maximum curvature points on these B-splines are considered as corners. Obviously such a location is not very precise. An alternative approach is the accurate location of such point of interest using a method proposed in (Deriche and Giraudon 1990). It has to be implemented in our system.

Another source of inaccuracy is the lens distorsion. We assume throughout this paper that the noise on the measures is uncorrelated, Gaussian and centered. Lens distorsion introduces correlated noise which is obviously not centered. Such noise has been estimated to a maximum of one pixel with our experimental set-up. It should be estimated very accurately and used for correcting the image measures in order to come closer to our noise hypotheses.

For the same problem, Faugeras (1992) and independently Hartley et al. (1992) provide an elegant linear projective reconstruction which heavily relies on the computation of the epipolar geometry and the associated fundamental matrix. The results we obtained with their method was much less accurate than the one we got with our approach; but as we were unable to reproduce their accuracy in the computation of the fundamental matrix, no comparison can be made right now. A common testbed will be set in the near future in order to be able to compare both approaches. Another advantage of the method proposed in this paper is that the solution is less sensitive to *bad* motions of the camera than the algebraic method (Faugeras 1992) due to the redundancy of the system of equations.

Acknowledgements

This work is partly supported by European Esprit BRA projects First and Second which are gratefully acknowledged. We are also 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 our colleague B. Boufama who provided the tracking algorithm.

References

Beyer, H.A. 1992. Accurate calibration of CCD cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 96–101.

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

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

Deriche, R., and Giraudon, G. 1990. Accurate corner detection: an analytical study. In *Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japan.*

Faugeras, O.D. 1992. 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.

Faugeras, O.D. 1993. *Three-Dimensional Computer Vision - A Geometric Viewpoint.* Artificial intelligence. M.I.T. Press, Cambridge, MA.

Faugeras, O.D., Luong, Q.T., and Maybank, S.J. 1992. 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.

Forsyth, D., Mundy, J.L., Zisserman, A., Coelho, C., Heller, A., and Rothwell, C. 1991. Invariant descriptors for 3D object recognition and pose. IEEE *Transactions on PAMI*, 13(10):971–991.

Hartley, R., Gupta, R., and Chang, T. 1992. R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764.

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

Lee, C.H., and Huang, T. 1990. Finding point correspondences and determining motion of a rigid object from two weak perspective views. *Computer Vision, Graphics and Image Processing*, 52:309–327.

Mohr, R., and Arbogast, E. 1990. It Can Be Done without Camera Calibration. Technical Report RR 805-I- IMAG 106 LIFIA, LIFIA–IMAG.

Mohr, R., Morin, L., Inglebert, C., and Quan, L. 1991. Geometric solutions to some 3D vision problems. In J.L. Crowley, E. Granum, and R. Storer, editors, *Integration and Control in Real Time Active Vision*, ESPRIT BRA Series. Springer-Verlag.

Mohr, R., Quan, L., Veillon, F., and Boufama, B. 1992. Relative 3D reconstruction using multiples uncalibrated images. Technical Report RT 84-I-IMAG LIFIA 12, LIFIA–IRIMAG.

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

Quan, L., and Mohr, R. 1992. Affine shape representation from motion through reference points. *Journal of Mathematical Imaging and Vision*, 1:145–151. also in IEEE Workshop on Visual Motion, New Jersey, pages 249–254, 1991.

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

Sparr, G. 1991. 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.

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