Algebraic Aspects of Reconstruction of Structured Scenes from One or More Views

Etienne Grossmann, Diego Ortin and José Santos-Victor
ISR, Instituto Superior Técnico
1049-001 Lisboa, Portugal
{etienne,dortin,jasv}@isr.ist.utl.pt

Abstract

This paper considers the problem of 3D reconstruction from 2D points in one or more images and auxiliary information about the corresponding 3D features : alignments, coplanarities, ratios of lengths or symmetries are known. Our first contribution is a necessary and sufficient criterion that indicates whether a dataset, or subsets thereof, defines a rigid reconstruction up to scale and translation. Another contribution is a reconstruction method for one or more images. We show that the observations impose linear constraints on the reconstruction. All the input data, possibly coming from many images, is summarized in a single linear system, whose solution yields the reconstruction. The criterion which indicates whether the solution is unique up to scale and translation is the rank of another linear system, called the "twin" system. Multiple objects whose relative scale can be arbitrarily chosen are identified. The reconstruction is obtained up to an affine transformation, or, if calibration is available, up to a Euclidean transformation.

1 Introduction

Reconstructing static scenes with some geometric structure has recently drawn a lot of attention [4, 11, 5]. By structure, it is meant that some sets of 3D points verify properties of coplanarity, alignment or symmetry or that some distance ratios are known. This situation is of practical importance because it is common in man-made scenes and because it may allow to obtain reconstruction from a single view. Possible applications can be found in urbanism (virtual models of existing or ancient buildings), leisure (models for virtual reality), real-estate (models of inside and outside of houses or appartments) etc.

The geometric information is given a-priori, as in [4, 11, 3, 2].

There are two main contributions in this paper. The first is a criterion that indicates whether a given dataset defines the reconstruction of one or more objects up to scale and translation. This criterion is calculated as the rank of a matrix introduced in Section 5.3 and is insensitive to noise. The second contribution is a method for obtaining a reconstruction. All the input data, which may come from many images, is "summarized" in a single linear system whose solution gives the reconstruction. In the presence of errors in the input image features, a least-squares solution is sought (Sec. 5.4).

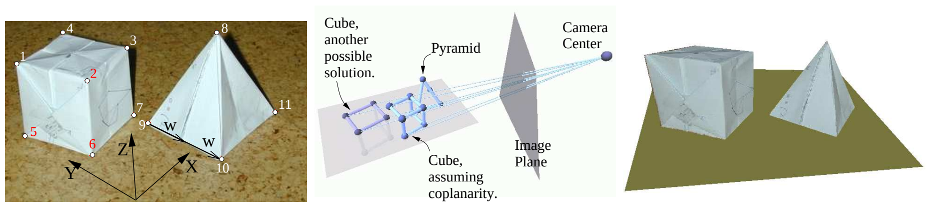

Figure 1: A simple 3D scene (left) and its reconstruction (right). If the bottom points were not known to be coplanar, the relative scale of pyramid and cube would be ambiguous (middle).

We use three "predominant" directions, which form the basis –not necessarily orthogonal [11, 12, 3]– in which reconstruction is obtained. The vanishing points[1] of these directions play a special role [3, 12, 2] and are assumed to have been estimated [10, 9] and given in the input data.

If the angles between the "predominant" 3D directions are known, partial calibration of the camera can be computed from the vanishing points [1, 12, 8, 2] and an Euclidean reconstruction will obtained. Otherwise, an "affine" reconstruction is obtained [6].

An example of input data is shown in Figure 1 (left). The predominant directions are the "x", "y" and "z" axes ("x" and "y" are aligned with the base of the pyramid). This dataset consists of 11 points and some auxiliary information : the 3D points (1-4), (5-7) and (9-11) lie in a horizontal plane. Also, points (1,5), (2,6) and (3,7) lie on three vertical lines, and (s)he has said that point 8 is midway, along the "y" axis between points 9 and 10. The input data is precisely defined in Section 2.2.

Published methods use auxiliary information that involve alignment, coplanarity [12, 2, 11], knowledge of world distances [3, 11] and angles. Additionally, the versatility of the method is increased by the use of known ratios of distances. In this way, some symmetry relations can be exploited. In all the presented experimental results (Sec. 6), some length ratios are known, whithout which the reconstructions could not be obtained.

An important contribution of the proposed method is that it determines whether there are many objects that can be scaled independently (Figure 1(middle)), and whether each one defines a rigid reconstruction (Prop. (1-3)). In the simplest case, if no length ratios or symmetries are known and if a single image is used, each object is always defined up to a scale factor (Prop. (2)). In the more general case, each object (Def. 1 and 2 and Prop. (1,3)) does not necessarily have a rigid solution, so it is indispensable to have a criterion that indicates the nature of the solution.

2 General conditions

In this section the notation is introduced and the input data is precisely defined.

[1]The vanishing point of a 3D direction is the image point in which intersect all projections of 3D lines parallel to that direction.

2.1 Notation

The three "predominant" 3D directions are called $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$. Points in $\mathbb{R}^3$ are always represented by their coordinates in the basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. We define $\mathbf{e}_1 = [1\,0\,0]^\top$, $\mathbf{e}_2 = [0\,1\,0]^\top$, $\mathbf{e}_3 = [0\,0\,1]^\top$. Lines in the image are represented by a 3×1 vector $\mathbf{l}$. The set of points contained in the line is $\left\{\mathbf{x} \in \mathbb{R}^2 | \left[\mathbf{x}^\top\, 1\right] \mathbf{l} = 0\right\}$.

2.2 Input Data

The input data consists in 2D points in the image(s) and *auxiliary information*, which indicates geometric properties of the corresponding 3D objects. The image points can be given in the pixel coordinates or, if calibration is known, in the Euclidean coordinate system associated with the camera.

2.2.1 Image features

1. Image points $\mathbf{x}_1, \ldots, \mathbf{x}_P$, projections of 3D points $\mathbf{X}_1, \ldots, \mathbf{X}_P$.
2. The vanishing points of the 3D directions $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. If $F > 1$ images are available, the vanishing points are called $\mathbf{r}_i^f$, $i \in \{1, 2, 3\}$. Each one is a 3×1 vector of homogeneous coordinates. These vectors [1] form the three first columns of the projection matrix (See Eq. (2)).

2.2.2 Auxiliary information

1. Knowledge that some observed 3D points belong to planes parallel to two of the canonical axes. Each plane is expressed as a list of (indices of) image features. For example, in Figure 1, the user would have specified that points (1-4) lie on a horizontal plane etc. Lines are formed by the intersection of two planes.
2. Information on ratios of distances taken along predominant directions. For example, the distance along the "y" axis from point 9 to 8 is equal to that from point 8 to 10. We call this information *"metric information"*.
3. If many images are available, one knows which image each 2D feature comes from.

Note that, when many images are given, planes may contain points observed in different images. A 3D point visible in many images can be "tracked" by defining a "x=constant", a "y=constant" and a "z=constant" plane that contain all the projections of that point. Also, metric information may relate points visible in different images and the related distances may be taken along different axes (See Sec. 6).

Auxiliary information is given through a text interface, but graphical interfaces can be imagined [4]. We do not know of automatic ways of obtaining auxilliary information from 2D points, much less from images, except in simple cases.

3 Use of auxiliary information

First, the set of distinct coordinates needed to describe the 3D data is determined and coordinates are related to image features. Then, using the metric information, the coordinates are expressed as a linear function of a subset of coordinates and of signed distances.

Consider, in the input data in Figure 1 (left), the line (10,11). Since this line is parallel to the "x" axis, the coordinates of these points are of the form $[C_1, C_2, C_3]$ and $[C_4, C_2, C_3]$ respectively. The second and third coordinates are identical. Then, considering that the line (9,10) is parallel to the "y" axis, the coordinates of point 9 are necessarily of the form $[C_1, C_5, C_3]$. By using all the user-supplied information, the set of distinct 3D coordinates $C_1, \ldots, C_N$ is identified, and one knows the correspondance between 2D points and 3D coordinates. This is easily implemented using basic set operations. The distinct coordinates are grouped in a vector $\mathbf{C} = [C_1, \ldots, C_N]^\top$.

Then, knowing that some distances, taken along coordinate axes, are equal, or have a known ratio u yields constraints of the form :

$$C_i - C_j = u\,(C_k - C_l)\,.$$

Defining the signed distance $W = C_k - C_l$, one gets

$$\begin{array}{rcl} C_k & = & C_l + W \\ C_i & = & C_j + uW\,, \end{array}$$

or, in matrix terms,

$$\begin{bmatrix} C_i \\ C_j \\ C_k \\ C_l \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \\ 1 & \\ & 1 \\ & 1 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} C_j \\ C_l \end{bmatrix}}_{\mathbf{C}_0} + \underbrace{\begin{bmatrix} u \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{U} \underbrace{[W]}_{\mathbf{W}}$$

The values of C_i, C_j, C_k and C_l are uniquely defined by C_j, C_l and W. Using all the user's input, $\mathbf{C}$ is represented as a linear function of a sub-vector $\mathbf{C}_0$ and of a vector signed of distances $\mathbf{W} = [W_1, \ldots, W_{P'}]^\top$:

$$\mathbf{C} = B\mathbf{C}_0 + U\mathbf{W}. \tag{1}$$

The choice of matrices B and U is not unique. One possible representations is chosen that minimizes the sum of lengths of $\mathbf{C}_0$ and $\mathbf{W}$ and also minimize the length of $\mathbf{W}$. If no metric information is used, B is the identity matrix and U is a zero-column matrix.

4 Use of image features

We now show how the observations impose linear constraints on $\mathbf{C}_0$ and $\mathbf{W}$. The observation are produced by a pinhole camera. The projection of a 3D point $\mathbf{X}$ with coordinates $[C_1, C_2, C_3]^\top$ (in basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$) to a 2D point $\mathbf{x} = [x_1, x_2]$ can be modeled as [8]:

$$\begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & -R\mathbf{T} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \tag{2}$$

for some $\lambda \in \mathbb{R}$. Here, $\mathbf{T}$ is the vector of (unknown) coordinates of the optical center.

If the projection $\left[\mathbf{x}^\top\ 1\right]^\top$ of a point with coordinates $C\mathbf{e}_i + C'\mathbf{e}_{i'} + C''\mathbf{e}_{i''}$ is observed, one may [8] build the 2D line $\mathbf{l}$ passing through that point and any one of the $\mathbf{r}_i$:

$$\mathbf{l} \sim \mathbf{r}_i \times \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}. \tag{3}$$

This 2D line is moreover the projection of the set of points with coordinates :

$$\{\mathbf{Y} \mid \exists\mu \in \mathbb{R},\ \mathbf{Y} = \mu\mathbf{e}_i + C'\mathbf{e}_{i'} + C''\mathbf{e}_{i''}\}\,. \tag{4}$$

The projection of a 3D point belonging to this line has the form :

$$\begin{aligned} \begin{bmatrix} \mathbf{y} \\ 1 \end{bmatrix} &= \lambda \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix} (\mu\mathbf{e}_i + C'\mathbf{e}_{i'} + C''\mathbf{e}_{i''}) - \lambda R\mathbf{T} \\ &= \lambda\,(\mu\mathbf{r}_i + C'\mathbf{r}_{i'} + C''\mathbf{r}_{i''} - R\mathbf{T}) \end{aligned} \tag{5}$$

If C', C'' and $\mathbf{T}$ are not known, but the user has located $\mathbf{l}$ in the image, one has a linear constraint on C', C'' and $\mathbf{T}$: any 2D point $\mathbf{y}$ in $\mathbf{l}$ verifies $\left[\mathbf{y}^\top\ 1\right]\mathbf{l} = 0$, so that, after expansion, one has :

$$0 = \mathbf{l}^\top\mathbf{r}_{i'}C' + \mathbf{l}^\top\mathbf{r}_{i''}C'' - \mathbf{l}^\top\mathbf{r}_{i'}T_{i'} - \mathbf{l}^\top\mathbf{r}_{i''}T_{i''} \tag{6}$$

This equation is a linear equation in the coordinates and in $\mathbf{T}$. One verifies that the three constraints given by each point (one constraint per vanishing point) form a system of rank two only.

5 Solutions to the reconstruction problem

The coordinates, distances and camera positions are solutions of a linear system obtained from the above-described constraints. This system may or may not define a unique reconstruction up to scale and translation. In the absence of noise in the observation, the ranks of certain subsystems indicate whether this is the case. In the presence of noise, the rank is altered, but a “twin” system may be built whose rank indicates whether the input data defines a reconstruction that is unique up to scale and translation.

5.1 Linear system

Concatenating Equations (6) obtained from the input data, one obtains a system of M equations :

$$[AB \mid AU] \begin{bmatrix} \mathbf{C}_0 \\ \mathbf{W} \end{bmatrix} + L \begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_F \end{bmatrix} = \mathbb{0}_{M,1} \ , \tag{7}$$

where A is the $M \times N$ matrix of coefficients that multiply the C_i and L, $M \times 3F$, multiplies the $\mathbf{T}_f$. We use the abbreviations $E = [AB \mid AU]$, and $\mathbf{V} = [\mathbf{C}_0; \mathbf{W}]$.

Row and column permutations may expose a block-diagonal structure in E. Each block corresponds to one “connected” object in the input data and we will use the terms “block” and “object” indifferently. Each block of E corresponds to a subset (defined by the columns of the block) of coordinates in $\mathbf{V}$, and a subset (defined by the rows) of 2D features. We will say that block E_p is *visible* in the images in which the 2D features appear.

It is assumed that $[E\,L]$ is itself single-block. If this is not the case, the data contains totally unrelated data sets. Each one can be treated separately, as described below.

After identifying, in E, blocks $E_1, \ldots, E_Q$ (if any) that are visible in one image only and grouping the remaining blocks (if any) in E', Eq. (7) becomes :

$$\begin{bmatrix} E_1 & & & \\ & \ddots & & \\ & & E_Q & \\ & & & E' \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_Q \\ \mathbf{V}' \end{bmatrix} + \begin{bmatrix} L_1 \\ \vdots \\ L_Q \\ L' \end{bmatrix} \begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_F \end{bmatrix} = \mathbb{0}_{M\times 1}, \tag{8}$$

Here, $\mathbf{V}$ has been split into $\mathbf{V}_1, \ldots, \mathbf{V}_Q$ and $\mathbf{V}'$. Each L_p is decomposed in $L_p = \left[L_p^1 \ldots L_p^F\right]$, where each L_p^f has three columns and multiplies one of the $\mathbf{T}_f$.

5.2 Nature of solutions : corank criteria

Def. 1 : We say that the reconstruction of block p, visible in image g, is *uniquely defined* up to a scale factor if and only if there is a $\mathbf{V}_p^*$ such that for all $\mathbf{V}_p$, $\mathbf{T}_g$ that solve $E_p\mathbf{V}_p + L_p^g\mathbf{T}_g = \mathbb{0}$, there is a scale factor λ_p such that $\mathbf{V}_p$ is of the form :

$$\mathbf{V}_p = \lambda_p \mathbf{V}_p^* + S_p \mathbf{T}_g. \tag{9}$$

Here S_p is defined in the following way : row number n of S_p, if it corresponds to a coordinate C_{0i}, taken along the j^{th} axis (1, 2 or 3 for “x”, “y” and “z”), is equal to $\mathbf{e}_j^\top$. Otherwise, if it corresponds to a distance w_i, it is equal to $\mathbb{0}_{1\times 3}$.

Equation (9) clearly displays the “scale and translation” interpretation, λ_p being the scale and $S_p\mathbf{T}_g$ the translation in the coordinates.

From now on, we assume that all the C_i are distinct and all the W_i are nonzero. One then has the following properties[2] :

Prop. 1 Reconstruction of block p is uniquely defined up to a scale factor if and only if E_p has corank equal to one.

Prop. 2 All blocks E_p defined without metric information have corank equal to one.

We now turn to the rest of the system :

Def. 2 We say that there is a *single rigid solution* to the system

$$E'\mathbf{V}' + L' \begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_F \end{bmatrix} = \mathbb{0} \tag{10}$$

if there exist some vectors $\Delta\mathbf{T}_2, \ldots, \Delta\mathbf{T}_F \in \mathbb{R}^3$ and $\mathbf{V}'^*$ such that for all $\mathbf{V}'$, $\mathbf{T}_1, \ldots, \mathbf{T}_F$ that solve (10), there exists a scale factor λ such that

$$\begin{aligned} \mathbf{T}_f &= \lambda\Delta\mathbf{T}_f + \mathbf{T}_1 \quad f \in \{2 \ldots F\} \text{ and} \\ \mathbf{V}' &= \lambda\mathbf{V}'^* + S\mathbf{T}_1 \end{aligned} \tag{11}$$

[2]Demonstrations are not given to save space. They appear in an article submitted to a journal.

Note that, in this definition, all the camera positions are uniquely defined by $\mathbf{T}_1$. Also, there is a single scale factor, even if E' can be block-diagonalized in more than one block.

The following property holds :

Prop. 3 There is a single rigid solution if and only if $[E'\, L']$ has corank four.

Properties (1-3) hold for all possible sizes of the E_p and E'.

5.3 Corank criteria in the presence of noise : twin matrices

The criteria given in Prop. (1-3) are valid when there are no errors in the input image features. In the presence of errors, the rank of submatrices of E and $[E\, L]$ is altered, so that the corank criteria proposed above cannot be used directly. However, a "twin matrix" may be built, that has the rank that $[E\, L]$ would have in the absence of noise. The corank criteria are used on the twin matrix.

The twin matrix has the same shape[3] as $[E\, L]$. Distinct coordinates $\mathbf{C}$ and camera positions $\mathbf{T}_f$ are generated randomly. The 3D lines Eq. (4) corresponding to these coordinates project in 2D lines :

$$\mathbf{l} \sim (C'' - T_{i''})\, \mathbf{s}_{i'} - (C' - T_{i'})\, \mathbf{s}_{i''}\ , \tag{12}$$

where $[\mathbf{s}_1\, \mathbf{s}_2\, \mathbf{s}_3] = R^{-\top}$. The twin system is built from these lines in the exact same way that $[E\, L]$ was built from the lines Eq. (3). One shows that noise in the vanishing points does not alter the rank of the twin matrix. In consequence, the twin matrix has the rank that $[E\, L]$ would have in the absence of noise. Using floating-point arithmetic, the rank of the twin matrix can be reliably computed [7], which guarantees that the corank criteria Prop. (1-3) can be computed from the twin matrix.

5.4 Computation of solutions

In this section, we show how to compute vectors $\mathbf{V}_p^*$, $\mathbf{V}^{'*}$ and $\Delta\mathbf{T}_f$ from the matrix $[E\, L]$ and how to obtain a particular solution to Eq. (8).

In the absence of noise and if $[E'\, L']$ has corank four, $\mathbf{V}^*$ and $\Delta\mathbf{T}_2, \ldots, \Delta\mathbf{T}_F$ verify :

$$E'\mathbf{V}^* + L'' \begin{bmatrix} \Delta\mathbf{T}_2 \\ \vdots \\ \Delta\mathbf{T}_F \end{bmatrix} = \mathbb{0}_{M\times 1},$$

where L'' is obtained by removing the first three columns in L' (which correspond to $\mathbf{T}_1$). Clearly, in the absence of noise, $[E'\, L'']$ has corank equal to one. In the presence of noise, this is not the case any more, but the singular vector [7] of $[E'\, L'']$ corresponding to the least singular value may be taken as an estimate of $[\mathbf{V}^*; \Delta\mathbf{T}_2; \ldots; \Delta\mathbf{T}_F]$. The $\mathbf{V}_p^*$ are likewise estimated by the singular vectors of E_p corresponding to the least singular value. A particular reconstruction is given by :

$$\mathbf{V}_p = \lambda_p \mathbf{V}_i^* + S_p \mathbf{T}_1 \,(p \in \{1 \ldots Q\}) \text{ and } \mathbf{T}_f = \Delta\mathbf{T}_f + \mathbf{T}_1 \,(f \in \{1 \ldots F\}) \tag{13}$$

for some $\lambda_1, \ldots, \lambda_Q$, $\lambda' \in \mathbb{R}$ and $\mathbf{T}_1 \in \mathbb{R}^3$.

[3] Zeros occur at the same places in both matrices.

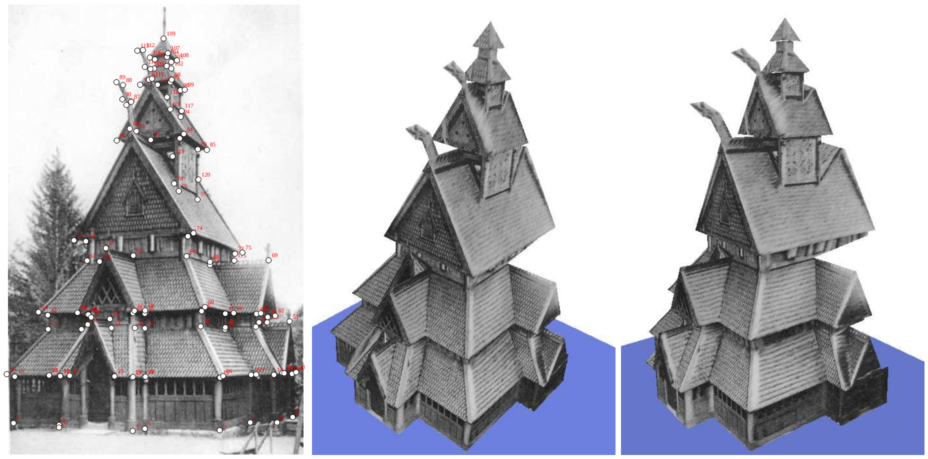

Figure 2: Original image (left) and two views of reconstruction.

5.5 Summary of the reconstruction algorithm

Identification: Identify distinct coordinates and distances and the correspondence between these and the observations.

Lines: For each 2D point $\mathbf{x}_i$, build the lines $\mathbf{l} \sim \left[\mathbf{x}_i^\top\ 1\right] \times \mathbf{r}_j$ between $\mathbf{x}_i$ and two at least of the vanishing points $\mathbf{r}_j$ in that image.

System: Using the lines obtained above, build the matrices E and L of the linear system. Block-diagonalize $[E\ L]$ and apply the following steps to each block.

Twin system: Generate random distinct coordinates and camera positions and the corresponding observations. From these, build the twin matrices of E and L.

Factorization: Factorize E in $E_1, \ldots, E_Q$ and E'.

Characterization: Determine the nature of the solutions from the rank of the twin matrices of each E_i and of $[E'\ L']$.

Reconstruction: Compute a solution to Eq. (8) as proposed in Section 5.4.

6 Experimental results

In this section, experimental results are presented. In all examples, the reconstruction basis is orthonormal.

Single image Figure 1 (right) shows the reconstruction obtained from the data described in the introduction, with the extra assumption that points (5-7 and 9-11) are coplanar. The vertical faces of the cube form an angle of 87 degree, which indicates that the Euclidean structure of the scene is well captured.

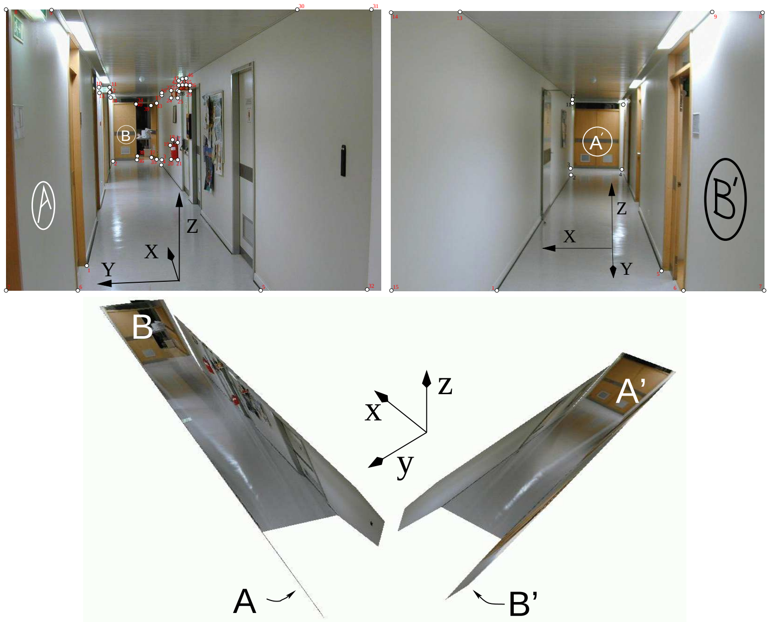

Figure 3: Two indoor images, looking ahead and to the right, and the reconstruction.

Figure 2 (left) shows an image with 122 points and (middle, right) two views of the corresponding reconstruction. The lengths of $\mathbf{C}$, $\mathbf{C}_0$ and $\mathbf{W}$ are 151, 102 and 26 respectively. Symmetry relations are needed to obtain a uniquely defined solution.

Multiple images Figure 3 shows two indoor images, taken from almost the same place, at approximately right angle. 61 points were identified, none being visible in both images. The two images are connected only by two horizontal planes (ceiling and floor) and one vertical plane : the left wall in the second image, which appears in the extreme right of the first image. Without metric information, the system would be single-block but without a single rigid solution : the second camera could be translated forward arbitrarily. However, by specifying that the two sides of the hall have equal lengths, one enforces the existence of a single rigid solution : the distance from the left wall in the first image (marked with an “A”) to the farthest door in the second image (marked “A-prime”) is equal to the distance from the right wall in the second image (marked “B-prime”) to the farthest door in the first image (marked “B”). Here, metric information relates features in different images.

7 Conclusions and future work

We have presented a method for 3D reconstruction from one or more views based on image features and auxiliary geometric information provided by the user. The main improvements in the proposed method are :

- A criterion, insensitive to noise, determines the nature of the solution, before the reconstruction is attempted.

- Many images may be processed at once rather than sequentially, as in [11].
- Some symmetry relations and, more generally, knowledge of distance ratios along the principal directions can be exploited.

The proposed method, which does not use special shapes, could add flexibility to a system such as [4], which requires object to fit in templates in its initial reconstruction phase.

If a probabilistic model of the error in the observations is given, maximum likelihood estimation could provide more precise and statistically characterized solutions. The present method could provide an initial estimate Maximum likelihood estimation is likely to be implemented by an iterative process requiring an initial solution, which could be provided by the presented method. Ongoing work also aims at extending the method to handle more than three predominant directions and auxiliary information relating the positions of the cameras.

Acknowledgements are due to grant PRAXIS / BD / 19594 / 99 and project Camera EC FTMP Network ERB FMRX-CT97-0127.

References

[1] B. Caprile and V. Torre. Using vanishing points for camera calibration. *IJCV*, 4:127–140, 1990.

[2] R. Cipolla and D.P. Robertson. 3D models of architectural scenes from uncalibrated images and vanishing points. In *In Proc. 10th IAPR*, pages 824–829, 1999.

[3] A. Criminisi, I.Reid, and A.Zisserman. Single view metrology. *IJCV*, 40(2):123–148, 2000.

[4] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs : A hybrid geometry- and image-based approach. In *SIGGRAPH '96*, pages 11–20, August 1996.

[5] A. Dick, P. Torr, and R. Cipolla. Automatic 3d modeling of architecture. In *Proc. BMVC2000*, volume 1, pages 372–381, 2000.

[6] O.D. Faugeras. Stratification of three-dimensional vision: projective, affine and metric representation. *J. Opt. Soc. Am.*, (A12):465–484, 1995.

[7] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. The Johns Hopkins University Press, Baltimore, MD, USA, third edition, 1996.

[8] Richard Hartley and Andrew Zissermann. *Multiple View Geometry in Computer Vision*. Cambridge, 2000.

[9] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architectural models from images. In *Proc. EuroGraphics*, volume 18, pages 39–50, September 1999.

[10] C. Rother. A new approach for vanishing point detection in architectural environments. In *BMVC2000*, pages 382–391, 2000.

[11] Heung-Yeung Shum, Mei Han, and Richard Szeliski. Interactive construction of 3d models from panoramic mosaics. In *CVPR*, pages 427–433, 1998.

[12] P.F. Sturm and S.J. Maybank. A method for interactive 3d reconstruction of piecewise planar objects from single views. In *Proc. BMVC*, pages 265–274, 1999.