Single and Multi-View Reconstruction of Structured Scenes

Etienne Grossmann, Diego Ortin and José Santos-Victor
ISR; IST - Torre Norte; Av. Rovisco Pais, 1; 1049-001 Lisboa; Portugal
{etienne,dortin,jasv}@isr.ist.utl.pt

Abstract

We present a method for reconstruction of structured scenes from one or more views, in which the user provides image points and geometric knowledge -coplanarity, ratios of distances, angles- about the corresponding 3D points. First, the geometric information is analyzed. Then vanishing points are estimated, from which camera calibration is obtained. Finally, an algebraic method gives the reconstruction.

Our algebraic reconstruction method improves the present state-of-the-art in many aspects : geometric knowledge includes not only planarity and alignment information, but also known ratios of lengths. The single and multiple-view cases are treated in the same way and the method detects whether the input data is sufficient to define a rigid reconstruction. We benchmark, using synthetic data, the various steps of the estimation process and show reconstructions obtained from real-world situations in which other methods would fail.

We also present a new method for maximum likelihood estimation of vanishing points.

1 Introduction

It has been shown [2, 6, 7, 5] that 3D reconstruction is achievable from a single image, provided that some geometric properties about the scene are known, as is possible in urban or indoors scenes. Possible applications include historical studies, urbanism, real-estate etc. In this article, we present a method for obtaining such reconstructions.

We consider structured scenes : there exist parallel lines and planes, some distances are equal or have known ratios. The directions that define these planes and lines play a special role and we call them "*dominant directions*".

The input could consist in one or more images, as in Fig. 1 (left). The user identifies the projections of 3D points to be estimated (white dots in the figure) and gives some geometric properties, such as :

- The five big white dots (at the right of the image) all belong to a "X-Z" plane. Other planes and lines have likewise been given.
- The slanted wall surface (bottom of walls) stick out by the same amount along the "X" and "Y" directions (distance "w_1" in the figure).
- The distances $\overline{ab}$, along the "X" axis is equal to the distance $\overline{ad}$ along the "Y" axis (distance "w_2"). Also, the distances $\overline{ac}$ along the "X" and "Y" axes are equal to $\cos(\pi/4)w_2$.

We will call *"geometric information"* all the coplanarity, alignment and distance ratio information given by the user. The input data is given a formal definition in Section 1.1.

The proposed method works in two steps : first, vanishing points[1] of the dominant directions are estimated and, if possible, the camera(s) is (are) calibrated. Then, the reconstruction is obtained by an algebraic method.

We outline here the algorithm, details being given in Sections 2-3. Each part of the algorithm is summarized at the end of the corresponding section.

Vanishing points Any number of dominant directions may be present. In Figure 1, there are five, labeled "X", "Y", "Z", "U" and "V". Maximum Likelihood estimates of the vanishing points are obtained (Section 2.1) under the assumption that the error on the observations are Gaussian, indepentent and identically distributed. Under the above assumptions, the likelihood function is the sum of the euclidean distance from the points to the lines that contain them, and pass through the vanishing point.

Calibration can be obtained if a right trihedron exists amongst the dominant directions [1]. Otherwise, an affine -rather than euclidean- reconstruction is obtained.

Projection matrices are obtained [1] from the vanishing points.

[1]The vanishing point of a 3D direction $\mathbf{v}$ is the unique image point in which all projections of 3D lines parallel to $\mathbf{v}$ intersect.

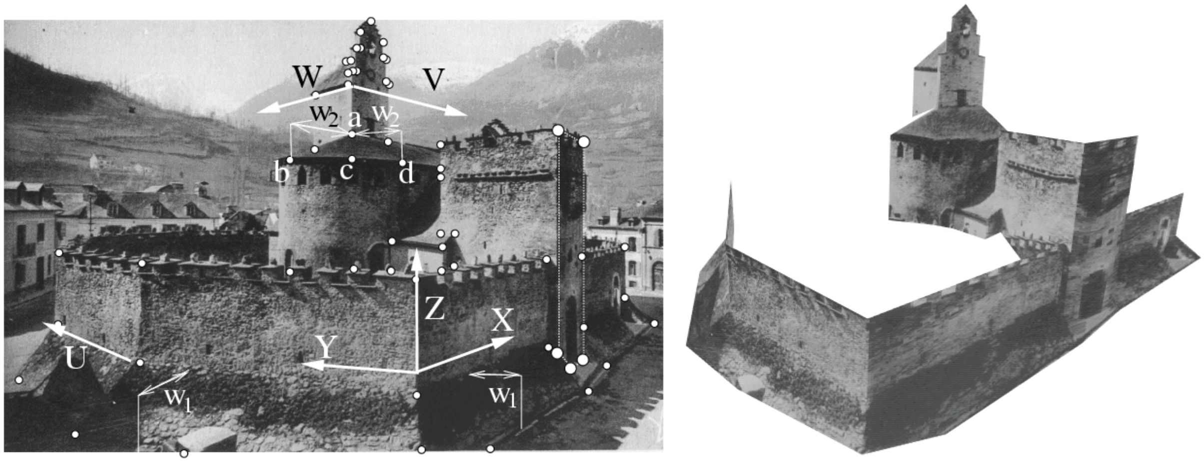

Figure 1. Left : Image with dominant directions (X,Y,Z,U,V,W) and 2D points. Right : Reconstruction.

The algebraic reconstruction method for obtaining reconstruction from 2D points and geometric information is the main contribution of this article. It's main characteristics are :

1. A criterion, insensitive to noise, is used to determine whether the input data defines a rigid reconstruction.

2. Known ratios of distances between parallel planes may be specified, which also allows to exploit symmetry in the scene.

3. All 3D points and camera position(s) are obtained simultaneously (unlike in [7, 6]). No reference plane [4] or shape template [2] is used.

4. Multiple and single-images cases are treated in the same manner. It is not necessary (although feasible) to track points across images to obtain reconstructions from many views.

5. The reconstruction always verifies *exactly* the geometric properties given by the user, unlike in [6, 7].

The algebraic reconstruction method first determines (Section 3.1) the linear constraints -referred to as "*geometric constraints*"- that are imposed on the coordinates of the 3D points by the geometric information. The set of feasible coordinates is a vector subspace, for which a basis is computed.

Then, the observed 2D points impose another set of linear constraints, "*observation constraints*", on the coordinates and on the camera positions (Section 3.2). In the presence of noise, a least-squares solution is sought.

In the noiseless case, and if the input data defines a rigid reconstruction, the geometric and the observation constraints are all simultaneously feasible, and the set of their solutions is a subspace of dimension four (Section 3.3). If this subspace is of greater dimension, this indicates that the input data is insufficient to define a rigid reconstruction. If there is noise in the observation, the rank is altered and this criterion cannot be used as-is. However, it is possible to generate a set of 3D points that verify the geometric constraints, and project them in the image plane, resulting in a *noiseless* set of observations; from these, one obtains noiseless observation constraints. We show that the original dataset defines a rigid reconstruction if and only if the set of coordinates that solve simultaneously the geometric and *noiseless* observation constraints is of dimension four. This criterion is insensitive to noise. In Section 3.3, we give a precise definition for a *"rigid reconstruction"* and the corresponding criteria.

In the rest of this section, the notations and assumptions are introduced and in Section 2 the estimation of vanishing points and calibration is presented. Section 3 describes the algebraic reconstruction method. Section 4 presents some experimental results : benchmarking of the algorithms is done using synthetic data, and results obtained from real-world images are shown.

1.1 Definition of the input data

The input data consists in 2D pixel coordinates $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^2$, localized in one of F images and *geometric information* representing known geometric properties of the 3D points and of the 3D dominant directions. This geometric information consists in :

- *Known angles and coplanarities* between dominant directions.

- *Planarity information*: Subsets of 2D points whose corresponding 3D points are known to belong to a 3D plane parallel to two of the dominant directions, which are also given. Alignement information can be represented by planarity information.

- *Metric information*: Pairs of pairs of parallel planes $(\mathcal{P}, \mathcal{P}')$ and $(\mathcal{Q}, \mathcal{Q}')$ and the knowledge of the ratio α of the distances between these planes. Each plane is defined by a 2D point $\mathbf{x}_m$ which it contains and by two directions $\mathbf{v}_i$ and $\mathbf{v}_j$ parallel to it.
- One knows what image each point $\mathbf{x}_n$ comes from.

1.2 Notations and assumptions

The 3D points and dominant directions will be written $\mathbf{X}_1, \ldots, \mathbf{X}_N$ and $\mathbf{v}_1, \ldots, \mathbf{v}_P$ respectively and identified with their coordinates in the basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. We assume the first three dominant directions are independent, so that they form a basis. If a right trihedron is given, we assume that it is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. We call $\mathbf{r}_1^f, \ldots, \mathbf{r}_P^f$ the vanishing points corresponding to the dominant directions in image number f ; all the $\mathbf{r}_i^f$ need not be observed.

Image lines are represented by a 3×1 vector $\mathbf{l}$. The set of 2D points contained in this line is $\left\{\mathbf{x} \in \mathbb{R}^2 | \left[\mathbf{x}^\top 1\right] \mathbf{l} = 0\right\}$.

The observations are obtained by perspective projection. Assuming $\mathbf{x}_m$ has been observed in image f, one has [3, 1] :

$$\begin{bmatrix} \mathbf{x}_m \\ 1 \end{bmatrix} = \lambda K_f \underbrace{\left[\mathbf{g}_1^f \, \mathbf{g}_2^f \, \mathbf{g}_3^f\right]}_{R_f} \left[\mathbf{X}_m - \mathbf{T}_f\right] + \begin{bmatrix} \varepsilon_m \\ 0 \end{bmatrix}, \quad (1)$$

where λ is a scale factor, K_f is the matrix of intrinsic parameters, $\mathbf{T}_f$ is the position of the camera in world coordinates. The error in the observations, the terms ε_m, are supposed to be independent, Gaussian and with covariance $\sigma^2 \left[\begin{smallmatrix} 1 & \\ & 1 \end{smallmatrix}\right]$, for some unknown σ.

2 Vanishing points and calibration

We now show how to estimate the vanishing points of some dominant directions and partial camera calibration.

2.1 Vanishing point estimation

First, the geometric information is examined to determine, for each direction, sets of 2D points that are the projections of 3D points on 3D lines parallel to that direction. A maximum likelihood estimate is computed for each vanishing point for which two or more lines are available.

A vanishing point $\mathbf{r}_i^f$ can be estimated if, in image f, the projections of at least two 3D lines, parallel to $\mathbf{v}_i$, are observed. The 2D points $\mathbf{x}_{m_1}, ..., \mathbf{x}_{m_\mu}$ are known to lie on the projection of a 3D line parallel to $\mathbf{v}_i$ if there exist two distinct planes, specified by the user, that contain the direction $\mathbf{v}_i$ and contain the points $m_1, ..., m_\mu$. Let $\mathcal{I}_1 = (m_1^1, ..., m_{\mu_1}^1), \ldots, \mathcal{I}_Q = (m_1^Q, ..., m_{\mu_Q}^Q)$ be the lists of indices of points in image f that belong to the projection of a 3D line parallel to $\mathbf{v}_i$. The maximum likelihood estimate of the vanishing point is the point $\mathbf{r}$ such that there exist lines $\mathbf{l}_1, ..., \mathbf{l}_Q$ passing through $\mathbf{r}$ that minimize the function :

$$(\mathbf{r}, \mathbf{l}_1, ..., \mathbf{l}_Q) \longrightarrow \sum_{j=1}^{Q} \underbrace{\sum_{k=1}^{\mu_j} d\left(\mathbf{l}_j, \mathbf{x}_{m_k^j}\right)^2}_{g_j(\mathbf{l}_j)},$$

where $d(\mathbf{l}, \mathbf{x})$ is the euclidean distance between the line $\mathbf{l}$ and the point $\mathbf{x}$. The search for the optimal $\mathbf{r}$ is greatly simplified by the fact that for all $\mathbf{r}$ and all j, the line $\mathbf{l}_j$ passing through $\mathbf{r}$ that minimizes $g_j(\mathbf{l}_j)$ is either the line $\hat{\mathbf{l}}_j$ that passes through $\mathbf{r}$ and $\bar{\mathbf{x}}_j = \sum_k \mathbf{x}_{m_k^j} / \mu_j$ (the centroid of the set of points), $\hat{\mathbf{l}}_j = \mathbf{r} \times \bar{\mathbf{x}}_j$, or the line $\hat{\mathbf{l}}_j'$ orthogonal to $\hat{\mathbf{l}}_j$ and passing through $\mathbf{r}$. One then has to minimize a function of $\mathbf{r}$ alone :

$$\mathbf{r} \longrightarrow \sum_{j=1}^{Q} \min \left\{ g_j\left(\hat{\mathbf{l}}_j\right), g_j\left(\hat{\mathbf{l}}_j'\right) \right\}.$$

This expression can be further simplified for easy computation. Its minimum is found using Nelder-Mead optimization [8] with multiple starting points.

2.2 Calibration

If the first three dominant directions form a right trihedron and K has the form

$$K = \begin{bmatrix} \rho & & u_0 \\ & \rho & v_0 \\ & & 1 \end{bmatrix}, \quad (2)$$

then it is well known [1] that ρ can be estimated from the vanishing points $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$, except in some pathological cases. In this article, we only estimate ρ and assume that $u_0 = v_0 = 0$.

2.3 Estimation of principal directions

It is clear that in the basis formed by the first three dominant directions, the coordinates of these directions are $\mathbf{v}_1 = [1\,0\,0]^\top$, $\mathbf{v}_2 = [0\,1\,0]^\top$ and $\mathbf{v}_3 = [0\,0\,1]^\top$. The coordinates of the other dominant direction, i.e. the vectors $\mathbf{v}_4, \ldots$ (if any) can then be estimated by :

$$\mathbf{v}_i = \left[\mathbf{r}_1^f \, \mathbf{r}_2^f \, \mathbf{r}_3^f\right]^{-1} \mathbf{r}_i^f.$$

2.4 Summary of algorithm (I)

1. Determine some vanishing pointby ML estimation and coplanarity constraints.
2. If a right trihedron is given, Estimate calibration matrices K_f. and left-multiply the observations $[\mathbf{x}_m\ 1]^\top$ the vanishing points $\mathbf{r}_i^f$ by K_f^{-1}.

3 Algebraic reconstruction method

In this section, the algebraic reconstruction method is presented. Three linear systems are considered : one is built from the geometric information and determines a linear subspace in which the reconstruction belongs. The second is built from the observed 2D features and defines the reconstruction. The third system has the same structure as the second and its corank determines whether the reconstruction is unique up to scale and translation.

3.1 Geometric constraints

We now show how to express the geometry information -coplanarity and metric information- as a linear system of equations on the coordinates of the 3D points. We build a basis for the set (a subspace) of coordinates that verify these equalities. Examining this basis, one checks whether the user provided coherent geometric information.

Planarity information If two points m, n are known to belong to a plane parallel to the directions i, j, one can easily show that the coordinates $\mathbf{X}_m, \mathbf{X}_n$ verify :

$$(\mathbf{v}_i \times \mathbf{v}_j)^\top \mathbf{X}_m - (\mathbf{v}_i \times \mathbf{v}_j)^\top \mathbf{X}_n = 0. \tag{3}$$

Metric information If points (m, n) (resp. (p, q)) lie on a pair of parallel planes $\mathcal{P}$, $\mathcal{P}'$ (resp. $\mathcal{Q}$, $\mathcal{Q}'$) with normal $\mathbf{v}$ (resp. $\mathbf{w}$) and one knows the ratio α of the distances from $\mathcal{P}$ to $\mathcal{P}'$ and from $\mathcal{Q}$ to $\mathcal{Q}'$, a linear constraint on the coordinates of points m, n, p and q can be found :

$$\mathbf{v}^\top (\mathbf{X}_m - \mathbf{X}_n) = \alpha \mathbf{w}^\top (\mathbf{X}_p - \mathbf{X}_q). \tag{4}$$

Note that the ratio of $\mathbf{v}^\top (\mathbf{X}_m - \mathbf{X}_n)$ over $\mathbf{w}^\top (\mathbf{X}_p - \mathbf{X}_q)$ is not invariant by affine transformations unless $\mathbf{v} = \mathbf{w}$. Thus it only makes sense to have $\mathbf{v} \neq \mathbf{w}$ if an Euclidean reconstruction is sought. The normals may be specified as the cross product of two dominant directions, or, if an Euclidean reconstruction is sought, by a dominant direction.

Using all the equations (3) and (4) provided by the geometric information, one gets a system of equations, the *geometric constraints* :

$$B\mathbf{X} = \mathbb{0}, \tag{5}$$

where $\mathbf{X} = \left[\mathbf{X}_1^\top, \ldots, \mathbf{X}_N^\top\right]^\top$ is the $3N \times 1$ vector holding all the point coordinates.

We call U a $3N \times M$ matrix whose columns form an orthonormal basis of the nullspace of B. All the solutions to Eq. (5) are of the form

$$\mathbf{X} = U\mathbf{V} \tag{6}$$

for some $\mathbf{V} \in \mathbb{R}^M$.

Examining U allows us to check the coherence of the geometric information : if B has full rank, or if some rows of U contain only zeros, the geometric constraints only have trivial solutions, most likely indicating an invalid geometric information.

3.2 Observations constraints

A linear system is now built from the 2D information, that constrains the coordinates of the reconstructed 3D points. In the presence of noise in the observations, this system may have no exact solution of the form (5) and the reconstruction is obtained as a least-squares solution.

Observation of a 2D point constrains the corresponding 3D point to lie on a 3D line. Not surprisingly, from each observed 2D point $\mathbf{x}_m$, one obtains two affine constraints on the coordinates $\mathbf{X}_m$.

For each observed point $\mathbf{x}$, projection of $\mathbf{X}$, and each vanishing point of the basis directions $\mathbf{g}_i$ (we omit the image number f), one can build the 2D line containing both points :

$$\mathbf{l} \sim \mathbf{g}_i \times \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}.$$

This line is the projection of the 3D line parallel to the i^{th} basis vector and passing through $\mathbf{x}$. Of course, one has $\left[\mathbf{x}^\top\ 1\right] \mathbf{l} = 0$, so that, using Eq. (1), one obtains :

$$\mathbf{l}^\top \mathbf{g}_{i'} X_i + \mathbf{l}^\top \mathbf{g}_{i''} X_{i''} - \mathbf{l}^\top \mathbf{g}_{i'} T_{i'} - \mathbf{l}^\top \mathbf{r}_{i''} T_{i''} \quad = 0 \ , \tag{7}$$

where i' and i'' are such that $\{i, i', i''\} = \{1, 2, 3\}$. The three linear equations Eq. (7) (one per vanishing point $\mathbf{g}_i$) obtained for each point form a system of rank two.

By concatenating the equations (7) obtained for each point, one obtains a linear system of *observation constraints :*

$$A\mathbf{X} + L\mathbf{T} = \mathbb{0}, \tag{8}$$

where $\mathbf{T} = \left[\mathbf{T}_1^\top \ldots \mathbf{T}_F^\top\right]^\top$ and A and L are $3N \times 3N$ and $3N \times 3F$ matrices holding the coefficients that multiply elements of $\mathbf{X}$ and $\mathbf{T}$, respectively. Since the geometric information constrains $\mathbf{X}$ to have the form of Eq. (6), Eq. (8) can be rewritten as :

$$AU\mathbf{V} + L\mathbf{T} = 0. \tag{9}$$

3.3 Nature of the solution

In this section we show how the rank of a certain matrix indicates whether the user provided data that defines a unique reconstruction up to scale and translation.

Definition: One says that a dataset *defines a rigid reconstruction* if and only if there exist vectors $\mathbf{X}^* = \left[\mathbf{X}_1^{*\top} \ldots \mathbf{X}_N^{*\top}\right]^\top \in \mathbb{R}^{3N}$ and $\Delta\mathbf{T}_2, \ldots, \Delta\mathbf{T}_F \in \mathbb{R}^3$ such that, for all $\mathbf{X}$ and $\mathbf{T}$ that verify Eqs. (9) , there is a scale factor $\lambda \in \mathbb{R}$ and a vector $\mathbf{T}_1 \in \mathbb{R}^3$ such that :

$$\begin{array}{rcll} \mathbf{X}_m & = & \lambda \mathbf{X}_m^* + \mathbf{T}_{\varphi_m} & (\forall m \in \{1...M\}) \\ \mathbf{T}_f & = & \mathbf{T}_1 + \lambda \Delta \mathbf{T}_f & (\forall f \in \{1...F\}) \end{array}$$

where $\varphi_m \in \{1...F\}$ is the index of the image in which $\mathbf{x}_m$ is observed.

Here, the $\mathbf{X}_m$ are "base" reconstructions, the $\Delta\mathbf{T}_f$ are displacements between camera positions, and λ, $\mathbf{T}_1$ are the arbitrarily chosen scale factor and position of the first camera.

If there is noise in the observations $\mathbf{x}_m$, Eq. (9) will have no exact solution. For a noisy dataset, one defines a rigid reconstruction as follows : assuming the observations are obtained by Eq. (1), one says that the dataset defines a rigid reconstruction if and only if the noiseless observations (Eq. (1) without the noise term ε_m) define a rigid reconstruction. Property B, below, says that, even without knowing the noiseless observations, it is possible to determine whether a dataset defines a rigid reconstruction.

Property A: In the absence of noise, there is a rigid reconstruction if and only if the matrix $[AU \,|\, L]$ has corank equal to four. We do not give here a proof of this statement.

If there is noise in the observations $\mathbf{x}_m$, the rank of the matrix $[AU \,|\, L]$ will be altered and one cannot use the criterion given above.

However, one can build matrices $\tilde{A}$ and $\tilde{L}$ such that there is a rigid reconstruction if and only if $\left[\tilde{A}U \,|\, \tilde{L}\right]$ has corank four. We call $\left[\tilde{A}U \,|\, \tilde{L}\right]$ the "twin matrix" of $[AU \,|\, L]$. This matrix is obtained in the following way : one generates randomly a vector $\tilde{\mathbf{V}} \in \mathbb{R}^M$ whose elements are all distinct, and distinct vectors $\tilde{\mathbf{T}}_1, ..., \tilde{\mathbf{T}}_F$. One defines $\tilde{\mathbf{X}} = U\tilde{\mathbf{V}}$ and then produces, using Eq. (1), *noiseless observations* $\tilde{\mathbf{x}}_m$, the projections of the 3D points $\tilde{\mathbf{X}}_m$. Finally, from these 2D points, one builds the matrices $\tilde{A}$ and $\tilde{L}$ in the same way that A and L were obtained from the $\mathbf{x}_m$.

Property B: There is a rigid reconstruction if and only if the "twin matrix" $\left[\tilde{A}U \,|\, \tilde{L}\right]$ has corank equal to four. We do not give here a proof of this statement.

Note that this criterion is not influenced *at all* by noise in the observations or in the vanishing points.

3.4 Computing a solution

We assume the twin matrix has corank four. The space of solutions to Eq. (9) would have dimension four in the absence of noise. In the presence of noise, we approximate this space by :

$$\left\{ \left[\begin{array}{c} \mathbf{V} \\ \mathbf{T} \end{array} \right] = H\mathbf{w} \right\}_{\mathbf{w}\in\mathbb{R}^4}$$

where the columns of H are the right singular vectors of $[AU \,|\, L]$ corresponding to the smallest singular values. It is often more convenient to represent the solution space by :

$$\left\{ \left[\begin{array}{c} \mathbf{X} \\ \mathbf{T} \end{array} \right] = \left[\begin{array}{c} UH_1 \\ H_2 \end{array} \right] \mathbf{w} \right\}_{\mathbf{w}\in\mathbb{R}^4},$$

where $H = \left[H_1^\top \; H_2^\top\right]^\top$ and H_1, H_2 have M and $3F$ rows respectively (corresponding to $\mathbf{V}$ and $\mathbf{T}$).

Obtaining a particular solution can be done by imposing four additional constraints to $\mathbf{X}$ and/or $\mathbf{T}$, for example by fixing the center of mass of the scene and its scale.

3.5 Summary of algorithm (II)

1. Build the matrix B from the vanishing points $\mathbf{r}_i^f$ and geometric information; compute a basis U for the nullspace of B, check that it contains only non-zero points (stop otherwise).
2. Build the matrices A and L from the vanishing points $\mathbf{g}_i^f$ and the observations $\mathbf{x}_m$.
3. Build the twin matrix $\left[\tilde{A}U \,|\, \tilde{L}\right]$. Verify that it has corank equal to four (stop otherwise).
4. Compute a basis for the space of solutions and a particular solution using other constraints.

4 Experimental results

4.1 Sensitivity to noise

We study the effect of noise on the algebraic reconstruction method -for Euclidean reconstruction, on the estimation of vanishing points and on the calibration process. Using synthetic data, with varying noise level, we study the error in the resulting reconstructions. A very wide range of noise levels is covered.

The "house" object shown in Figure 2 (left), consisting of ten points, is used. It's coordinates are all within the interval $[-1, 1]$. Five directions, "Z", "X", "Y", "U" and "V" are present, the last four being coplanar. Nine planes are identified. The object is rotated randomly and observed by perspective projection, the matrix K in Eq. (1) is of the form (2) where ρ, the focal length, is generated randomly with a uniform density probability function in $[1.5,\, 2]$ -not an uncommon value if pixel coordinates have been normalized to $[-1,\, 1]$. The principal point $[u_0,\, v_0]$ is likewise taken to be uniformly distributed in $[-0.05,\, 0.05]$.

Noise is added to these observations with amplitudes varying from 25dB (6% error : the standard deviation of the error is 0.06 times that of the observations) to 60dB (0.1% error). In real-world situations, we believe the noise levels are in the range 0.3-1% (40-50dB).

The vanishing points are estimated and calibration is obtained from the dominant directions "X", "Y" and "Z". Then, the algebraic reconstruction method is used to obtain the reconstruction, in three different cases :

1. Using the maximum likelihood vanishing points and focal length estimated as in [1].

2. Using the maximum likelihood vanishing points and the true calibration matrix K.

3. Using the true vanishing points and calibration matrix.

This experiment was repeated 50 times. The error between the true and estimated parameters are measured.

Vanishing points and calibration Figure 2 (middle) shows the mean absolute error, measured in degrees in the estimated vanishing points. The smooth curve shows the error in the vanishing points estimated from two lines ("U" and "V" axes) and the dashed curve is for vanishing points estimated from four ("Y","Z") or five ("X") lines.

Algebraic reconstruction The smooth curve in Figure 2 (right) shows the mean error of the reconstruction algorithm when the true vanishing points and calibration are given; the dashed curve is for estimated vanishing points and known calibration, and the dotted curve is for estimated vanishing points and calibration.

For values that are common in practice, between 0.3% and 1% the error level is seen to be very reasonable. For higher noise levels, the error increases approximately linearly, showing the robustness of the algorithm.

4.2 Real-world data

Figure 1 shows on the left a real-world image and on the right the reconstruction obtained from it. Fifty nine (59) points and 50 planes and nine ratios of lengths were given. The matrix U has 57 columns. If one reprojects the estimated 3D points in the image, the noise level with respect to the observed points is 29.5dB.

Figure 3 shows two indoor images taken from approximately the same point, but in more-or-less perpendicular directions. The input consists in 61 points, 35 planes and one known ratio of lengths : the distance from the point marked "A" in the first image to that marked "A-prime" in the second image is equal to that from point "B" (first image) to point "B-prime" second image. No 3D point is visible in both images. The error level of reprojected points with respect to observed points is 42.9dB.

Figure 4 (left,middle) shows two outdoors image with some overlap. Seventy-two (24 in the first image, 48 in the second) points and 21 planes were identified. In order to obtain a rigid reconstruction, it is necessary to use metric information : one assumes that the spikes on the left and middle wall stick out by the same amount. Without this knowledge, the relative scale of the spikes and of the rest of the scene would be undetermined. Figure 4 (bottom) shows the reconstruction. The error level of reprojected points with respect to observed points is 46.6dB.

5 Conclusions and future work

We have presented a method for 3D reconstruction from one or more views that checks whether the input data is coherent and sufficient to define a rigid reconstruction. It could e.g. add versatility to an easy-to-use interactive reconstruction system such as [2, 5, 7].

It allows reconstruction from many images (like [6], but one does not need to track points across images [Sec. 4.2]). We have shown that using metric information increases the versatility : we believe none of the presented real-world results could be obtained using other published methods.

Also, we have presented and benchmarked a method for computing maximum likelihood estimates of vanishing points.

There are many prospects for future development : one could include constraints on $\mathbf{T}$ in the geometric information; maximum likelihood estimation of the reconstruction has also been implemented in the presented framework.

Credits : This work has been supported by grants 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] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs : A hybrid geometry- and image-based approach. Technical Report CSD/96-893, UCB, January 1996.

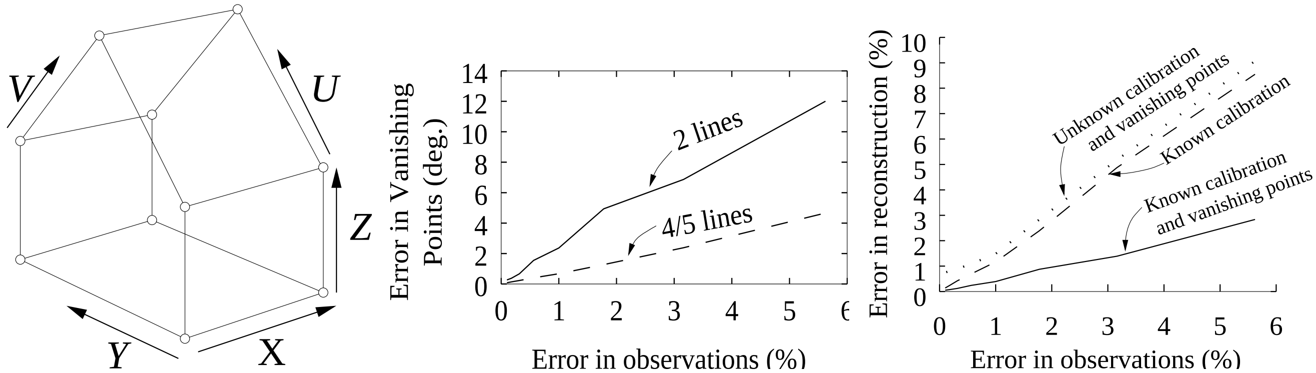

Figure 2. Left : synthetic object. Middle : (Mean absolute) Error (in degrees) in the estimated vanishing points, using 2 lines (smooth curve) or 4-5 lines (dashed curve). Right : Error in observations vs. error in reconstruction.

Figure 3. Left: Two indoor images. No 2D point is tracked between them Right : reconstruction.

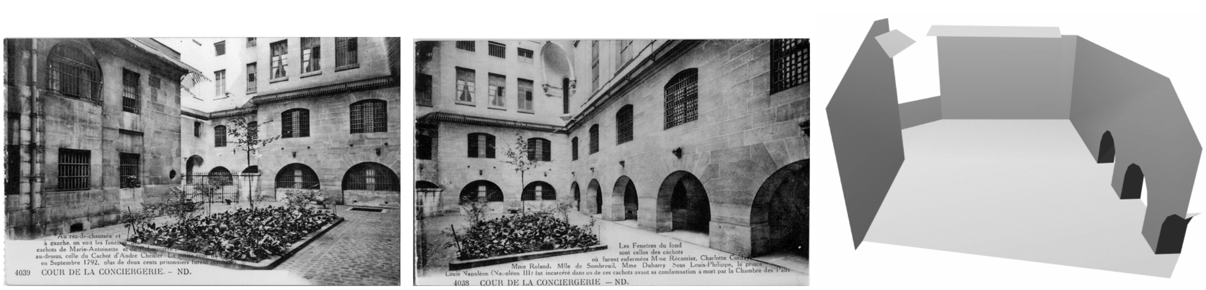

Figure 4. Left: Two outdoor images. Right: Untextured reconstruction

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

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

[5] D. Robertson and R. Cipolla. An interactive system for constraint-based modelling. In *BMVC2000*, volume 2, pages 536–545, 2000.

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

[7] 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.

[8] William T. Vetterling William H. Press, Saul A. Teutolsky and Brian P. Flannery. *Numerical Recipes, The Art of Scientific Computing, 2nd edition*. Cambridge University Press, 1992.