Using geometric clues in 3D reconstruction from a single view

Etienne Grossmann and José Santos Victor

(etienne|jasv)@isr.ist.utl.pt,

www.isr.ist.utl.pt/~(etienne|jasv)

July 22, 2002

Abstract

We outline a method for performing single-view reconstruction. There exist many possible approaches, using different techniques and assumption, reaching various degrees of automatism and we focus on the reconstruction of environments that are rich in planes, alignements, symmetries, orthogonalities and other forms of geometrical regularity. Finding these 3D properties in images is best done -as of today- by the human visual system and we assume that a human operator provides this geometric information. Also, the operator has chosen from the image the 2D points whose 3D position will be estimated. Given this 2D information and some geometric information about the corresponding 3D points, we determine whether the 3D shape is defined uniquely and how to reconstruct it. The proposed method expresses the geometric constraints as a system of linear constraints and transforms the reconstruction problem into a linear algebra problem, with the benefit that properties of the reconstruction problem can be deduced from those of the linear problem.

Keywords: single-view 3D reconstruction, geometric constraints, vanishing points.

1 Introduction

We consider the problem of obtaining a 3D reconstruction from 2D points localized in a single image and from some geometric information concerning the corresponding 3D points. Planarity, colinearity, known angles and other geometric properties, provided by a "user" will be used to disambiguate the scene and, if possible, obtain a unique reconstruction. We will address the important questions that arise when solving this problem while omitting the proofs. Also, we do not provide a sensitivity analysis of the reconstruction method, which is best done using other techniques. This document focuses on finding out whether the input data is coherent and sufficient and provides means to compute a reconstruction using tools of numerical linear algebra.

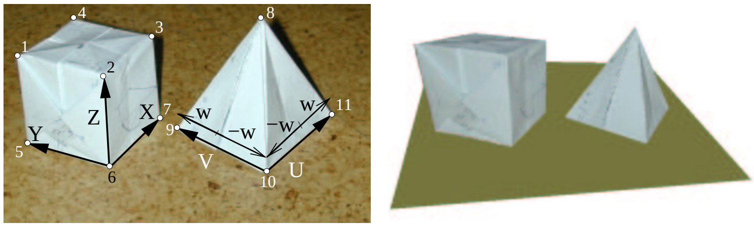

Figure 1: **Left:** Simple dataset consisting in 2D points, numbered 1 to 11 and in user-supplied geometric information about the scene. Planarities : points $\{1,2,3,4\}$ and $\{5,6,7,9,10,11\}$ belong to distinct horizontal plane; points $\{1,2,5,6\}$ belong to a plane orthogonal to the "X" axis, and the sets points $\{2,3,6,7\}$, $\{9,10\}$ and $\{10,11\}$ belong to planes orthogonal to the axes "Y", "U" and "V" , respectively. Also, the user indicated that point 8 is mid-way along the "V" (resp. "U") axis between point 9 and 10 (resp. 10 and 11). Finaly, it is known that the triplet of axes {X, Y, Z} and {U, V, Z} form right trehidra. **Right:** Reconstruction obtained from that information, decorated with facets and texture.

To begin with an example, Figure 1 (left) and its caption illustrate the various kinds of geometric information that are used in the reconstruction method. Although very simple, it illustrates all the kinds of information that are given by the user. The desired output consists in the positions of the 3D points and of the camera, represented by coordinates in an orthogonal basis attached to the camera. Some camera calibration parameters are also estimated, together with some intermediate quantities, such as the directions "U" , "V",..., "Z" .

Having defined the problem, we introduce the mathematical symbols that represent the quantities of interest. The coordinates of 2D points that were identified in the image are 2×1 vectors $\mathbf{x}_1, ..., \mathbf{x}_N$, which have been normalized so that they lie in $[-1, 1] \times [-1, 1]$[1]. The coordinates of the corresponding 3D points are written $\mathbf{X}_1, ..., \mathbf{X}_N \in \mathbb{R}^3$. These points are observed by perspective projection [8, 6] :

$$\left[\begin{array}{c} \mathbf{x}_m \\ 1 \end{array}\right] = \lambda_m \underbrace{\left[\begin{array}{ccc} f & 0 & u \\ 0 & f & v \\ 0 & 0 & 1 \end{array}\right]}_{K} (\mathbf{X}_m - \mathbf{T}) + \left[\begin{array}{c} \varepsilon_m \\ 0 \end{array}\right], \tag{1}$$

where $\lambda_m = 1/(X_{m3} - T_3)$ is the inverse of the "depth", K is the matrix of intrinsic parameters and $\mathbf{T} = [T_1\, T_2\, T_3]^\top$ is the position of the camera in world coordinates. The error term ε_m is due to optical imperfections and finite resolution of the camera and to small errors commited when the 2D points were localized in the image.

[1] For example, pixel coordinates $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2] \in [0, w] \times [0, h]$ are transformed into $\mathbf{x} = [\tilde{x}_1 - w/2, \tilde{x}_2 - h/2] / \max\{w, h\}$.

The reconstruction method described in this paper can be roughly cut in three parts. In the first, like in other published methods [9, 11, 2], the orientation and calibration of the camera are estimated (Section 2). Knowing these quantities, we transform the geometric information into a system of *linear* constraints on the coordinates of the 3D points (Section 3). Finally, the 2D observations are used to further constrain the 3D points and obtain the reconstruction (Section 4). Also, it is shown in Section 5 how to verify whether the user provided sufficient information. Conclusions are given in Section 6.

2 Estimating plane orientations and camera calibration

The calibration of the camera and orientation of the considered planes can be estimated from vanishing points using the well-known method of Caprile and Torre [1], which we briefly overview.

The vanishing points themselves are obtained by identifying, thanks to the geometric information provided by the user, sets of 3D points that belong to segments lines to some directions of interest. For example, in Figure 1, points $\{1,2\}$, $\{3,4\}$ and $\{5,6\}$ belong to three distinct lines parallel to the "X" direction. The vanishing point corresponding to the "X" direction can e.g. be found as the 2D point that minimizes the sum of squared distances to the 2D lines defined by points $\{1,2\}$, $\{3,4\}$ and $\{5,6\}$. In order avoid problems with points at infinity, vanishing points will be represented by a vector $\mathbf{g} \in \mathbb{R}^3$ that furthermore verifies $\|\mathbf{g}\| = 1$. We assume from now on that the vanishing points $\mathbf{g}_1, ..., \mathbf{g}_5$ of the directions "U" ,..., "Z" have been estimated.

In short, there are two important ideas to the calibration method of [1]. First, one notes that a 3D direction can be identified with its vanishing point (Figure 2). That is, an observed vanishing point $\mathbf{g}_i$ differs from the corresponding 3D direction $\mathbf{v}_i$ only by the transformation induced by the calibration parameters :

$$\mathbf{g}_i = \lambda K \mathbf{v}_i \text{ for some } \lambda \in \mathbb{R}.$$

Second, if three vanishing points are known, of three 3D directions that form a right trihedron (i.e. each pair forms a $\pi/2$ angle), then any deviation from orthogonality in the observed vanishing points is only due to the matrix of calibration parameters. As a consequence, finding the calibration is equivalent to finding the 3×3 matrix K of the form of Eq. (1) that "rectifies" the vanishing points so that they become two-by-two orthogonal. Except in some not-so-rare critical configurations (we do not enter this subject here), it is possible to obtain K by solving numerically a system of nonlinear equations.

In what follows, we assume that the first three vanishing points $\mathbf{g}_1$, $\mathbf{g}_2$, $\mathbf{g}_3$ correspond to three mutually orthogonal directions and that the calibration matrix K has been estimated, e.g. by the method of [1]. Once the camera is calibrated, estimates of the directions $\mathbf{v}_i$ are computed by :

$$\mathbf{v}_i = \lambda^{-1} K^{-1} \mathbf{g}_i. \qquad (2)$$

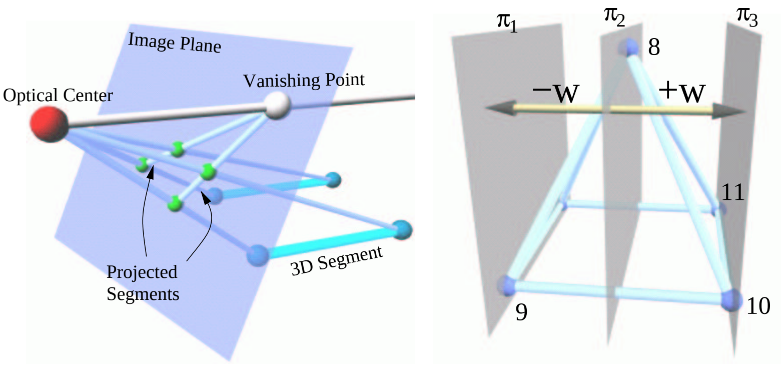

Figure 2: **Left:**A 3D direction can be identified with the coordinates of its vanishing point in a calibrated image because the vector defined by the optical center and the vanishing point is parallel to the 3D direction.
Right: The symmetry in the pyrammid in Figure 1 can be expressed as an equality of distances between pairs of parallel planes. For example, the signed distances between the pairs of planes (π_1, π_2) and (π_2, π_3) are opposite.

3 Using geometric clues

Having identified and estimated the directions $\mathbf{v}_i$ that are relevant in the considered scene, the geometric information provided by the user can be converted into linear equations that constrain the coordinates of the 3D points.

Planarity In order to express that points $\mathbf{X}_m$ and $\mathbf{X}_n$ belong to a plane with normal $\mathbf{v}_i$, it is equivalent to say that :

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

Since the $\mathbf{v}_i$ have been estimated, one has a linear constraint on $\mathbf{X}_m$ and $\mathbf{X}_n$. If a plane contains N' points, e.g. $\mathbf{X}_1, ..., \mathbf{X}_{N'}$, then $N' - 1$ equations of this kind can be found, for example by writing Eq. (3) with $(m, n) = (1, 2), ..., (m, n) = (N' - 1, N')$. It can be verified that these equations are independent, that they imply that the points are coplanar and that, reciprocously, if the points are contained in a plane with normal $\mathbf{v}_i$, then these equations are verified. In other terms, these $N' - 1$ equations are equivalent to saying that the 3D points belong to a plane with normal $\mathbf{v}_i$.

Ratio of distances between pairs of parallel planes We now show how some symmetries and other types geometric properties can also be expressed by a linear equation. Going back to the example in Figure 1, the fact that «point 8 lies midway along the "V" axis between points 9 and 10» can equivalently be expressed by the equation :

$$\mathbf{v}_5^\top (\mathbf{X}_8 - \mathbf{X}_9) = -\mathbf{v}_5^\top (\mathbf{X}_8 - \mathbf{X}_{10}), \tag{4}$$

where it is assumed that $\mathbf{v}_5$ defines the "V" axis. Figure 2 (right) shows the geometric interpretation of this equation. A more general way of expressing this type of geometric constraints is given in the following equation :

$$\mathbf{v}_i^\top (\mathbf{X}_m - \mathbf{X}_n) + \alpha \mathbf{v}_j^\top (\mathbf{X}_p - \mathbf{X}_q) = 0. \tag{5}$$

Here, the distances are not necessarily equal, but have a known ratio α. Also, by taking $\mathbf{v}_i \neq \mathbf{v}_j$, the distances need not be taken along the same direction.

Origin of coordinates Finally, it is convenient to fix the origin of the referential, for example so that it coincides with the center of mass of the reconstruction. This can be expressed by the following three linear equations :

$$[I_3, \ldots, I_3]\, \mathbf{X} = \mathbb{O}_{3\times 1}, \tag{6}$$

where $\mathbb{O}_{3\times 1}$ represents a 3×1 matrix of zeros and I_3 is the 3×3 identity matrix.

The constraints and their solution Joining together all the geometric information, converted in linear equations Eq. (1), Eq. (5) and Eq. (6), one gets a single system :

$$B\mathbf{X} = \mathbb{O}_{M\times 1}, \tag{7}$$

where $\mathbf{X} = \left[\mathbf{X}_1^\top, ..., \mathbf{X}_N^\top\right]^\top$ is a $3N \times 1$ vector holding all the coordinates of the 3D points and B is a $M \times 3N$ matrix holding the coefficients of the equations. This equation characterizes the sets of configurations of N 3D points that verify the geometric constraints, for the considered directions $\mathbf{v}_1, ..., \mathbf{v}_D$.

In other terms, the nullspace [10] of B is the set of all vectors of coordinates $\mathbf{X} \in R^{3N}$ that verify all the geometric constraints supplied by the user. If U is a $3N \times Q$ matrix whose columns form an orthonormal basis of the nullspace of B, then any vector $\mathbf{X}$ that verifies all the geometric constraints can be written

$$\mathbf{X} = U\mathbf{V} \tag{8}$$

for some vector $\mathbf{V} \in \mathbb{R}^Q$. This equation thus parameterizes the set (linear subspace of $\mathbb{R}^{3N}$) of collections of 3D points that verify all the geometric constraints supplied by the user.

It is to some extent possible to determine whether the user provided coherent geometric information by examining B and U. First, if B has full rank, then its nullspace is $\{\mathbb{O}\} \subset \mathbb{R}^{3N}$, which indicates that a configuration of 3D points verifies all the geometric constraints only if all the points are equal and most likely indicates that the user made an error in the specifications of the geometric constraints. If B is not full rank, and if two triplets of rows of U, numbered $\{3m-2, 3m-1, 3m\}$ and $\{3n-2, 3n-1, 3n\}$ are equal, then the geometric constraints given by the user imply that the two 3D points $\mathbf{X}_m$ and $\mathbf{X}_n$ are equal. Since these 3D points correspond to distinct 2D observations $\mathbf{x}_m$ and $\mathbf{x}_n$, this most likely indicates a mistake by the user. These tests for checking the coherence of the geometric information will be completed in Section 5 by a test on its sufficiency.

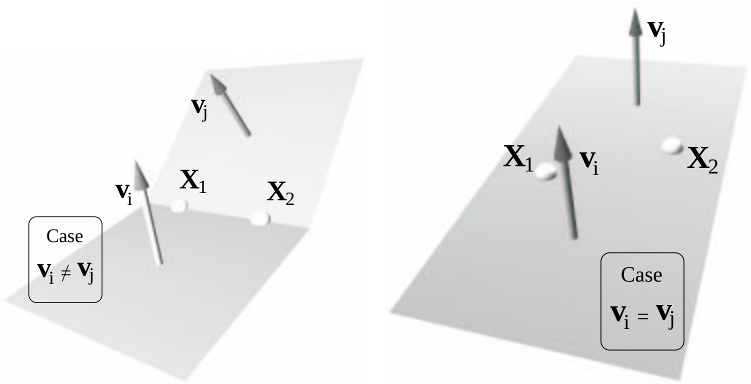

Figure 3: **Left :** If the dominant directions $\mathbf{v}_i$ and $\mathbf{v}_j$ are different, then points $\mathbf{X}_1$ and $\mathbf{X}_2$ that belong simultaneously to a plane with normal $\mathbf{v}_i$ and a plane with normal $\mathbf{v}_j$ are constrained to belong to a line. **Right :** If the dominant directions $\mathbf{v}_i$ and $\mathbf{v}_j$ are equal, the same planarity properties only constrain the points $\mathbf{X}_1$ and $\mathbf{X}_2$ belong to a plane, and there is one extra degree of freedom.

An important assumption should now be made, that the rank of B (i.e. the number of columns of U) does not change when the directions $\mathbf{v}_1, ..., \mathbf{v}_D$ are subject to small changes. This ensures that the nature of the shape is not changed because of the errors in the estimated $\mathbf{v}_i$.

To illustrate this point (figure 3), one may e.g. consider two points $\mathbf{X}_1$, $\mathbf{X}_2$ constrained to lie on two planes with normals $\mathbf{v}_i$ and $\mathbf{v}_j$ respectively. If these vectors are not colinear, then the points are constrained to lie on a line and this configuration is characterized by 4 parameters : the position of the line (two parameters) and their abscissa on the line (2×1 parameters). If we built the matrix B corresponding to these planarity constraints, one would obtain a 2×6 matrix of rank two and the corresponding matrix U is 6×4.

Now, if $\mathbf{v}_i$ and $\mathbf{v}_j$ are colinear (figure 3, right), then the points are only constrained to lie on a plane and the configuration is characterized by five parameters, one for the plane and 2×2 to specify the positions of the points in the plane. In terms of the framework introduced above, the matrix B now has rank one and U is 6×5.

There exist other more suble situations in which, for some configurations of the $\mathbf{v}_i$, the rank of B and the size of U vary when the $\mathbf{v}_i$ are subject to small perturbations. In these cases, the matrices B obtained from the "true" vectors $\mathbf{v}_i$ and from those estimated by Equ. (2) are likely to differ in rank and the resulting matrices U will differ in size. Our assumption thus says that we are not in one of these cases. In practice, such cases are rare and our assumption does not constitute a limitation.

4 Using 2D observations

We now use the 2D observations $\mathbf{x}_1, ..., \mathbf{x}_N$ to add extra linear constraints on $\mathbf{X}$ and $\mathbf{T}$ and obtain the reconstruction of the scene. Recalling that the collinearity of two 3D vectors can be expressed by saying that their cross product is zero, one sees from Eq. (1) that :

$$\left[\begin{array}{c}\mathbf{x}_m\\1\end{array}\right] \times K\left(\mathbf{X}_m - \mathbf{T}\right) = -\underbrace{\left[\begin{array}{c}\mathbf{x}_m\\1\end{array}\right] \times \left[\begin{array}{c}\varepsilon_m\\0\end{array}\right]}_{\text{Small error term}}. \qquad (9)$$

Each 3D point $\mathbf{X}_m$ (and $\mathbf{T}$) is thus constrained by three linear equations that form a system of rank two - the remaining indeterminacy is the detph. By joining together the equations of this type obtained from all observations, and ignoring the error term in the right hand side, one obtains a system (of rank $2N$) :

$$\underbrace{\left[\begin{array}{ccc}S_1 & & \\ & \ddots & \\ & & S_N\end{array}\right]}_{A}\mathbf{X} + \underbrace{\left[\begin{array}{c}-S_1\\ \vdots \\ -S_N\end{array}\right]}_{L}\mathbf{T} = A\mathbf{X} + L\mathbf{T} = \mathbb{0}_{3N\times 1}, \qquad (10)$$

where each 3×3 block S_m has the form :

$$\left[\begin{array}{ccc}0 & -1 & x_{m2}\\ 1 & 0 & -x_{m1}\\ -x_{m2} & x_{m1} & 0\end{array}\right]K.$$

In order to limit the search for solutions of Eq. (10) to the configurations of 3D points that verify all the geometric constraints, one may replace Eq. (8) in Eq. (10), and obtain :

$$AU\mathbf{V} - L\mathbf{T} = [AU \,|\, L]\left[\begin{array}{c}\mathbf{V}\\ \mathbf{T}\end{array}\right] = \mathbb{0}_{3N\times 1}. \qquad (11)$$

In the absence of error in the observations and *if the matrix* $[AU\,|\,L]$ *has corank equal to one*, then this system has a unique (up to scale) solution $\mathbf{V}^* \in \mathbb{R}^Q$, $\mathbf{T}^* \in \mathbb{R}^3$, and the reconstructed 3D points and camera position are have the form :

$$\mathbf{X} = \mu \underbrace{U\mathbf{V}^*}_{\mathbf{X}^*} \text{ and } \mathbf{T} = \mu\mathbf{T}^*,$$

where μ is an arbirtrary scale factor.

It should be noted that if $[AU\,|\,L]$ has corank two or more (has two or more singular values equal to zero), then the $[\mathbf{V};\, \mathbf{T}]$ that solve Eq. (11) form a space of dimension two or more, and thus the solution is not unique up to scale. This occurs when the geometric information provided by the user is not sufficient, and it is important to detect these cases. We assume until the next section that the user provided sufficient information and show how to obtain the reconstruction in the presence of noise.

In the presence of noise, the singular values of $[AU \mid L]$ are usually altered so that this matrix has full rank and Eq. (11) does not have an exact solution. In that case, one finds a solution "in the total least-squares sense" [7] by setting $\left[\mathbf{V}^{*\top}\, \mathbf{T}^{*\top}\right]^\top$ to be the singular vector of $[AU \mid L]$ corresponding to the least singular value. The resulting reconstruction $\mathbf{X} = U\mathbf{V}^*$ still verifies exactly (up to machine precision) the geometric constraints.

5 Caveat : are the geometric clues sufficient?

In this section, we show how to determine whether the user provided sufficient information to define uniquely the reconstruction. As was noted, in the absence of noise, a unique reconstruction is defined only if the matrix $[AU \mid L]$ has corank equal to one. In the presence of noise, the rank of this matrix is altered, so that it is not possible to use it directly to determine whether sufficient information was provided.

One solution would be to use a threshold on the smallest singular values of this matrix in order to decide whether its corank should be considered to be one or two. This approach requires studying the perturbation of the singular values of the matrix and the resulting test would have a nonzero probability of failure.

For these reasons, we adopt a different path, which allows to determine whether the input data is sufficient to define a unique reconstruction in a way that is totally insensitive to noise in the observations. This method results from the observation that a matrix "analogous" to $[AU \mid L]$ can be built without using the noisy observations and whose corank is one if and only if the user provided sufficient geometric information.

First, one notes that it is possible to build a collection $\mathbf{X}'$ of 3D points that verify the geometric constraints specified by the user, by randomly choosing a vector $\mathbf{V}' \in \mathbb{R}^M$ and taking $\mathbf{X}' = U\mathbf{V}'$. Then, a random camera position $\mathbf{T}'$ is chosen and the perspective projections $\mathbf{x}'_1, ..., \mathbf{x}'_N$ are computed using Eq. (1). From these *noiseless* 2D points, matrices A' and L' are built in the exact same way that A and L were built from the $\mathbf{x}_i$.

The resulting $[A'U \mid L']$ has many properties that $[AU \mid L]$ would have in the absence of noise, because it is obtained from the noiseless observation of a collection of 3D points that verify the same geometric properties. In particular, if the geometric information is sufficient, then the set of 3D collections that verify both the geometric constraints (i.e. are of the form $\mathbf{X} = U\mathbf{V}$ for some $\mathbf{V}$) and project to the 2D points $\mathbf{x}'_1, \ldots, \mathbf{x}'_N$ is of the form $\{\mu\mathbf{X}' \mid \mu \in \mathbb{R}\}$. Indeed, the contrary would indicate that the reconstruction is not defined uniquely, up to a scale factor by the geometric information and 2D points. Reciprocously, if the data is not sufficient, then the equation

$$[A'U \,|\, L'] \begin{bmatrix} \mathbf{V} \\ \mathbf{T} \end{bmatrix} = \mathbb{O}_{3N\times 1} \qquad (12)$$

will have solutions that are not of the form

$$\begin{bmatrix} \mathbf{V} \\ \mathbf{T} \end{bmatrix} = \mu \begin{bmatrix} \mathbf{V}' \\ \mathbf{T}' \end{bmatrix},$$

and matrix $[A'U \,|\, L']$ has corank two or more (since $\left[\mathbf{V}'^\top \mathbf{T}'^\top\right]^\top$ belongs to its nullspace, as well as another vector that is not colinear to it). Thus, the dimension of the nullspace of $[A'U \,|\, L']$ -its corank- indicates whether the input data is sufficient or not. Since this matrix can be computed without using the noisy observations, it provides a test for the sufficiency of the geometric information that is totally insensitive of the noise in the observations.

6 Conclusions, extensions

We have thus shown how the problem of reconstruction of scenes from 2D points and geometric information can be treated with tools of linear algebra, with the benefit that the coherence and sufficiency of the geometric information can be determined.

The proposed method can be (and has been, in other publications [3, 4]) extended in many ways, for example to treat many images simultaneously and to reconstruct many disconnected objects in the scene, whose relative scale cannot be determined.

Our study focuses on expressing the reconstruction problem in terms of linear systems, but it does not consider the sensitivity of the reconstruction with respect to noise in the observations. In order to treat the problem of reconstruction in such a way that the precision is known, a different approach and different tools should be used. For example, the framework of maximum likelihood estimation is used in [5]. Because that method is iterative, it should be initialized, for example with the method proposed in the present study.

References

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

[2] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. In *Proc. ICCV*, pages 434–441, Corfu, Greece, 1999.

[3] Etienne Grossmann, Diego Ortin, and José Santos-Victor. Algebraic aspects of reconstruction of structured scenes from one or more views. In *Proc. BMVC*, volume 2, pages 633–642, 2001.

[4] Etienne Grossmann, Diego Ortin, and José Santos-Victor. Single and multi-view reconstruction of structured scenes. In *Proc. ACCV*, pages 228–234, 2002.

[5] Etienne Grossmann and José Santos-Victor. Maximum likelihood 3d reconstruction from one or more images under geometric constraints. In *Proc. BMVC*, 2002.

[6] R. Hartley and A. Zissermann. *Multiple View Geometry in Computer Vision.* Cambridge University Press, 2000.

[7] S. Van Huffel and Joos Vandewalle. *The Total Least-Squares Problem : Computational Aspects and Analysis.* SIAM, 1991.

[8] R. Mohr and B. Triggs. *Projective Geometry for Image Analysis*, chapter http:// www.dai.ed.ac.uk/ CVonline/ LOCAL_COPIES/ MOHR_TRIGGS/ node9.html. CVonline, 1996.

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

[10] G. W. Stewart. *Introduction to matrix computations.* Academic Press, 1973.

[11] 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.