



Maximum Likelihood 3D Reconstruction from One or More Images under Geometric Constraints

Etienne Grossmann* and José Santos-Victor
{etienne|jasv}@isr.ist.utl.pt

Abstract

We address the 3D reconstruction of scenes in which some planarity, collinearity, symmetry and other geometric properties are known à-priori. Our main contribution is a reconstruction method that has advantages of both constraint-based and model-based methods.

Like in the former, the reconstructed object needs not be an assemblage of predefined shapes. Like in the latter, the reconstruction is a maximum likelihood estimate and its precision can be estimated. Moreover, we improve on other constraint-based methods by using symmetry and other forms of regularity in the scene, and by working indifferently with one or more images.

A second contribution is a method for parameterising a configuration of 3D points subject to geometric constraints. Using this parameterisation, the maximum likelihood reconstruction is obtained by solving an unconstrained optimisation problem.

Another contribution lies in validating experimentally the assumption under which the maximum likelihood estimator was defined, namely, that the errors in hand-identified 2D points behave approximately like identically distributed independent Gaussian random variables. With this assumption validated, benchmarking is performed on synthetic data and the precision obtained on real-world data is shown. These experiments show that the maximum likelihood estimator is well-behaved and give insight on the precision obtained in real-world situations.

1 Introduction

When reconstructing a 3D scene from observed 2D points, knowing à-priori geometric information facilitates the task [2, 6, 1] and in some cases allows to obtain a reconstruction from a single image [5, 13, 14]. These advantages allow many new applications such as rapid acquisition of virtual models of buildings [6, 5, 13, 4, 14], reconstruction from paintings and precise measurements [5].

We consider the reconstruction of structured scenes from 2D points in one or more images, when a user gives beforehand geometric information about the scene, such as planarity, collinearity, symmetry etc. This problem has been addressed in the past, either

*This work was supported by PRAXIS XXI Grant BD / 19594 / 99



in model-based approaches [12, 6, 17], in which predefined geometric shapes are fitted to image features or in constraint-based [2, 1, 15, 13, 5, 14, 17] methods, in which geometric constraints are imposed on the reconstructed 3D points. The presented method shares aspects of both approaches.

Like in model-based methods, the shape of the reconstruction depends on some parameters. However, we do not parameterise an assemblage of predefined shapes, so that the user has more freedom in choosing how to represent the scene. Because it is independent from the reconstruction process, the presented parameterisation technique could be used in a model-based reconstruction method [6] and increase its flexibility.

Like in constraint-based methods, the geometric information is given in the form of constraints on the 3D points. We improve over other methods by using extended geometric constraints, such as symmetry and other forms of geometric regularity and by using one or more images indifferently. Because we refine by least-squares fitting a reconstruction obtained by an algebraic method¹, there is a superficial resemblance with some previously published single-view methods [13, 14]. However, these authors do not detail how optimisation is done, while we treat this question in detail. By transforming the reconstruction problem into one of unconstrained optimisation, the dimensionality of the problem is reduced, classical optimisation tools can be used and it is moreover possible to estimate the precision of the reconstruction [10].

Determining whether the user gives information that is consistent and sufficient to define a unique reconstruction is an important issue that is not addressed here ([8]) because it is quite distinct from the maximum likelihood method described in the present article.

There are two principal steps in the method, first to express the geometric information as equality constraints on the estimated quantities (Section 3) and then to transform the reconstruction problem into one of unconstrained optimisation. Due to the geometric constraints, a constrained optimisation problem is first obtained. Then, like in [1], we use our differentiable parameterisation of the set of 3D points subject to the constraints (Section 4) and obtain an unconstrained optimisation problem. Unlike in [1], advanced polynomial manipulations are not needed and a result of differential matrix calculus [8] is used instead.

The reconstruction is obtained by least-squares, which is the right thing to do when the errors in the 2D observations are Gaussian, an assumption that is often made, but is rarely verified experimentally. We are only aware of [16, Appendix B] where such a verification is done, in conditions different from ours², so that it is important (Section 5.1) to perform an experimental verification. Then, the estimator is benchmarked on synthetic data and the precision obtained on real-world examples is shown. Finally further discussion and conclusions are given in Section 6.

2 Problem statement

We now define formally the problem of reconstruction. We assume that F images are available, in which N 2D observations $\mathbf{x}_1, \dots, \mathbf{x}_N$ are the perspective projections of 3D

¹We do not discuss here this algebraic method whose details have been published elsewhere.

²In [16], automatically tracked points are considered, while we consider hand identified-points which are not necessarily tracked and further subject to error due to inexactitude in geometric constraints.



points $\mathbf{X}_1, \dots, \mathbf{X}_N$. If \mathbf{x}_m is observed in image number f ($1 \leq f \leq F$), one has [11] :

$$\begin{bmatrix} \mathbf{x}_m \\ 1 \end{bmatrix} = \lambda_m^f K R_f (\mathbf{X}_m - \mathbf{T}_f) + \begin{bmatrix} \varepsilon_m \\ 0 \end{bmatrix}, \quad (1)$$

where λ_m^f is the ‘‘inverse depth’’, \mathbf{T}_f is the position of the camera in world coordinates, R_f is the rotation matrix that relates the camera frame to the world frame and K is the matrix of intrinsic parameters (constant in all images). We assume for now that only the focal length is unknown, although more intrinsic parameters could be estimated [3, 14].

The errors in the observations are represented by the ε_m , which are independent, Gaussian random variables with covariance $\text{cov}(\varepsilon_m) = \sigma^2 I_2$, where I_2 is the 2×2 identity matrix and σ is unknown. Other forms of the covariance matrix could be considered, e.g. that returned by an automatic feature tracker, but we assume that the scaled identity models faithfully the error caused by the human operator and the inexactitude in the geometric model. The validity of this assumption is verified experimentally in Section 5.1.

To facilitate later discussion, we define the ‘‘observation function’’ \mathcal{P} that associates to a 3D point \mathbf{X} , a rotation matrix R , a camera position \mathbf{T} and camera calibration matrix K , the perspective projection of \mathbf{X} ,

$$\mathbf{x} = \mathcal{P}(\mathbf{X}, R, \mathbf{T}, K) = \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} K R (\mathbf{X} - \mathbf{T}), \quad (2)$$

For convenience, the parameters \mathbf{X}_m , R_f , \mathbf{T}_f and K are grouped in a single entity :

$$\mathcal{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N, R_1, \dots, R_F, \mathbf{T}_1, \dots, \mathbf{T}_F, K), \quad (3)$$

and another function called \mathcal{P} is defined :

$$\mathcal{P}(\mathcal{X}) = \begin{bmatrix} \mathcal{P}(\mathbf{X}_1, R_{f_1}, \mathbf{T}_{f_1}, K) \\ \vdots \\ \mathcal{P}(\mathbf{X}_N, R_{f_N}, \mathbf{T}_{f_N}, K) \end{bmatrix} \in \mathbb{R}^{2N}. \quad (4)$$

where f_m is the number of the image in which \mathbf{x}_m is observed. This function $\mathcal{P}()$ is disambiguated from that in Eq. (2) by the context.

Hence, our goal consists in estimating the \mathbf{X}_m , R_f , \mathbf{T}_f and K , based on noisy observations, \mathbf{x}_m . In addition, some geometric information provided by the user will be used, as is detailed in the next section.

3 Geometric constraints

We now explain how the geometric information given by the user is transformed into equality constraints on the coordinates of the 3D points in the scene.

Planarity, collinearity and parallelism. If two points \mathbf{X}_m and \mathbf{X}_n are known to belong to a plane with unit normal \mathbf{v} , the following equation must hold :

$$\mathbf{v}^\top (\mathbf{X}_m - \mathbf{X}_n) = 0. \quad (5)$$

Collinearities are represented combining two planarities, by using the property that two points lie on a segment parallel to a 3D direction \mathbf{v} , if and only if they belong to two planes that have (non collinear) normals orthogonal to \mathbf{v} .

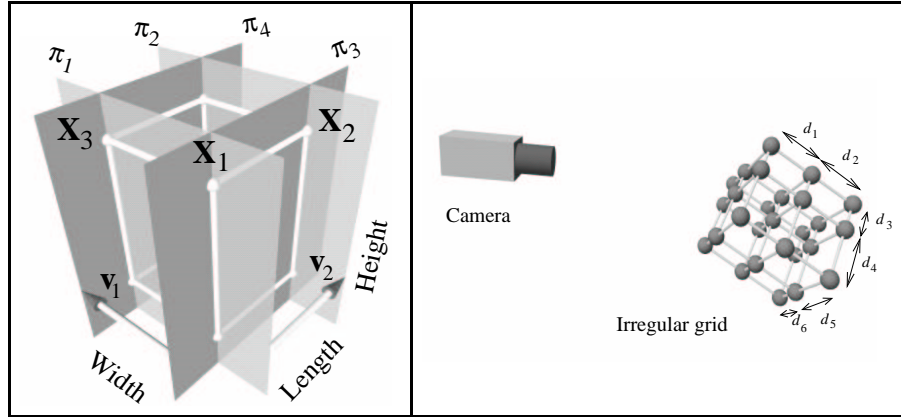


Figure 1: **Left** : The fact that the parallelepiped is as wide as long is expressed by the equation $\mathbf{v}_2^\top (\mathbf{X}_1 - \mathbf{X}_2) = \mathbf{v}_1^\top (\mathbf{X}_1 - \mathbf{X}_3)$. **Right** : Synthetic setup used for benchmarking the estimator.

Ratios of distances between pairs of parallel planes. Figure 1 (left) shows a parallelepiped whose width and length are equal. This equality is expressed, if \mathbf{v}_1 and \mathbf{v}_2 are the normals of the vertical planes, by :

$$\mathbf{v}_2^\top (\mathbf{X}_1 - \mathbf{X}_2) = \mathbf{v}_1^\top (\mathbf{X}_1 - \mathbf{X}_3) .$$

More generally, the distances between the planes need not be equal, but only have a known ratio $\alpha \in \mathbb{R}$. For arbitrary points $\mathbf{X}_m, \mathbf{X}_n, \mathbf{X}_p$ and \mathbf{X}_q , this type of constraint takes the form :

$$\mathbf{v}_i^\top (\mathbf{X}_m - \mathbf{X}_n) - \alpha \mathbf{v}_j^\top (\mathbf{X}_p - \mathbf{X}_q) = 0. \quad (6)$$

All the known equations resulting from planarities and known ratios of distances are assembled into an expression involving all the 3D points and geometric constraints. The coordinates of the points are joined in a single $3N \times 1$ vector $\mathbf{X} = [\mathbf{X}_1; \dots; \mathbf{X}_N]$ and a single matrix equation is obtained

$$B(\mathbf{v}_1, \dots, \mathbf{v}_D) \mathbf{X} = \mathbf{0}_{M \times 1}, \quad (7)$$

where M is the total number of constraints and the distinct plane normals are $\mathbf{v}_1, \dots, \mathbf{v}_D$. The notation $B(\mathbf{v}_1, \dots, \mathbf{v}_D)$ emphasises that this matrix depends only on the \mathbf{v}_i .

If $U(\mathbf{v}_1, \dots, \mathbf{v}_D)$ is a $3N \times P$ matrix whose columns form an orthonormal basis of the nullspace of $B(\mathbf{v}_1, \dots, \mathbf{v}_D)$, then all solutions to Eq. (7) can be written

$$\mathbf{X} = U(\mathbf{v}_1, \dots, \mathbf{v}_D) \mathbf{V}, \quad (8)$$

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

Constraints on the 3D directions Typically, some information on the \mathbf{v}_i is also available, such as known angles and planarities. Usually, three of the directions form a right trihedron, and in the present work, we take $\mathbf{v}_1 = [1, 0, 0]^\top, \mathbf{v}_2 = [0, 1, 0]^\top$ and $\mathbf{v}_3 = [0, 0, 1]^\top$. Section 4 explains how constraints on the other \mathbf{v}_i are used within the estimation process.

