



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*Using Geometric Constraints Through  
Parallelepipeds for Calibration and 3D Modelling*

Marta Wilczkowiak — Peter Sturm — Edmond Boyer

N° 5055

Novembre 2003

THÈME 3

A large blue rectangle occupies the lower half of the page. Overlaid on it is a large, light grey 'R' on the left. To the right of the 'R', the words 'Rapport de recherche' are written in a white serif font. A horizontal grey brushstroke is positioned below the text.

*Rapport  
de recherche*





## Using Geometric Constraints Through Parallelepipeds for Calibration and 3D Modelling

Marta Wilczkowiak , Peter Sturm , Edmond Boyer

Thème 3 — Interaction homme-machine,  
images, données, connaissances  
Projet MOVI

Rapport de recherche n° 5055 — Novembre 2003 — 51 pages

**Abstract:** This paper concerns incorporation of geometric information in the camera calibration and 3D modeling. Using the geometric constraints enables stabler results and allows to perform tasks with fewer images. Our approach is motivated and developed within a framework of semi-automatic 3D modeling, where the user defines geometric primitives and constraints between them. It is based on the observation that constraints such as coplanarity, parallelism or orthogonality, are often embedded intuitively in parallelepipeds. Moreover, parallelepipeds are easy to delineate by a user, and are well adapted to model the main structure of e.g. architectural scenes. In this paper, first a duality that exists between the shape of a parallelepiped and the intrinsic parameters of a camera is described. Then, a factorization-based algorithm exploiting this relation is developed. Using the images of the parallelepipeds, it allows to simultaneously calibrate cameras, to recover shapes of parallelepipeds and to estimate the relative pose of all entities. Dealing with a well constrained three-dimensional structure makes it possible to overcome the common problems of the factorization methods: missing data and unknown scale factors. The reconstruction obtained this way is of the affine character. To remove the affine ambiguity, all the available metric information: constraints on parallelepipeds' edge lengths and angles, as well as the usual self-calibration constraints on cameras can be used simultaneously. The proposed algorithm is completed by a study of the singular cases of the calibration method. Also a method for the reconstruction of scene primitives that are not modeled by parallelepipeds is introduced. The method is validated by various experimental results for real and simulated scenes, for cases where a single or several views are available.

**Key-words:** camera calibration, geometrical constraints, parallelepipeds, factorization, 3D scene analysis

## Comment utiliser les parallélépipèdes pour calibrer des images et construire des modèles 3D

**Résumé :** Ce rapport traite de l'utilisation de contraintes géométriques dans les processus de calibrage de caméras et de modélisation 3D à partir d'images. Ces processus sont en effet sensibles aux bruits et les contraintes géométriques, introduites par l'utilisateur, permettent d'en stabiliser les résultats même lorsque peu d'images sont considérées. Nous nous focalisons, dans ce document, sur les contraintes géométriques associées à un type de primitive particulier, les parallélépipèdes. Ces primitives permettent de caractériser des contraintes de coplanarité, de parallélisme et d'orthogonalité. De plus, les parallélépipèdes sont facilement identifiables par l'utilisateur dans les images et sont bien adaptés à la modélisation de bâtiments ou d'environnements architecturaux. Nous mettons en évidence, dans ce rapport, le fait que l'image d'un parallélépipède de caractéristiques connues caractérise entièrement la transformation de projection dans l'image. Il y a en fait dualité entre les caractéristiques du parallélépipède et celles de la caméra au travers de la transformation de projection. Nous développons ensuite un algorithme de factorisation qui exploite cette relation de dualité et permet de calibrer et de reconstruire simultanément l'ensemble des caméras et des parallélépipèdes définis par l'utilisateur dans un jeu d'images. L'intérêt des parallélépipèdes est de résoudre les difficultés classiques des méthodes de factorisation appliquées aux points : les données manquantes et les facteurs d'échelle inconnus. La reconstruction obtenue de cette manière est de type affine. Pour lever l'ambiguïté, l'ensemble des informations métriques sur les parallélépipèdes et sur les caméras peut être utilisées simultanément. Nous complétons la présentation de l'algorithme par une étude des configurations singulières pour le calibrage. Une méthode de reconstruction des primitives qui ne sont pas modélisées par les parallélépipèdes est aussi présentée. Les performances de l'algorithme sont démontrées au travers d'expériences sur des données synthétiques et réelles, pour des jeux d'une seule ou de plusieurs images.

**Mots-clés :** calibrage de caméras, contraintes géométriques, parallélépipèdes, factorisation, analyse 3D

## Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Preliminaries</b>	<b>9</b>
2.1	Camera parameterization . . . . .	9
2.2	Parallelepiped parameterization . . . . .	9
<b>3</b>	<b>Projections of Parallelepipeds</b>	<b>12</b>
3.1	One Parallelepiped in A Single View . . . . .	12
3.2	$n$ Parallelepipeds in $m$ Views . . . . .	14
<b>4</b>	<b>Estimating Intrinsic and Orientation Parameters by Factorization</b>	<b>16</b>
4.1	Missing Data . . . . .	16
4.2	Recovery of Scale Factors . . . . .	17
4.3	Factorization . . . . .	18
4.4	Disambiguating the Factorization . . . . .	19
4.4.1	Using Knowledge on Camera Intrinsic . . . . .	20
4.4.2	Prior Information on Parallelepipeds . . . . .	21
4.5	Complete Algorithm . . . . .	22
4.6	Minimal cases for the linear calibration . . . . .	22
<b>5</b>	<b>Singularities</b>	<b>23</b>
5.1	One parallelepiped in a single view . . . . .	23
5.2	One parallelepiped in multiple views . . . . .	26
<b>6</b>	<b>Estimating Position and Size</b>	<b>28</b>
<b>7</b>	<b>3D Reconstruction</b>	<b>29</b>
<b>8</b>	<b>Experimental results</b>	<b>30</b>
8.1	Synthetic Scenes . . . . .	30
8.1.1	A minimal case: 1 camera and 1 parallelepiped . . . . .	30
8.1.2	Scenario with two cameras . . . . .	32
8.2	Results on real scenes . . . . .	35
8.2.1	Bedroom scene . . . . .	36
8.2.2	Notre-Dame square scene . . . . .	37
8.2.3	Opposite viewpoint scene . . . . .	39
8.2.4	Kio towers . . . . .	39
<b>9</b>	<b>Conclusion</b>	<b>40</b>

<b>A Proof of the singularities for the calibration of cameras with known skew and principal point</b>	<b>42</b>
A.1 1 parallelepiped seen by 1 camera . . . . .	44
A.2 1 parallelepiped seen in 2 cameras . . . . .	45

## 1 Introduction

Efficient 3D modeling from images is one of the most challenging issues in computer vision. The tremendous research effort made to develop feasible methods in this domain has proved that recovering 3D structures from 2D images is a difficult and often under-constrained problem. Several reasons account for that, including the fundamental fact that without any prior information on cameras, or on the scene to recover, a metric reconstruction is not possible at all [13, 19]. This is why knowledge on the acquisition process, or on the scene, is required. A number of approaches have been proposed to exploit various prior information, both on camera and scene parameters. Such prior information do not only solve the projective ambiguity in the reconstruction but do also usually stabilize the sensitive reconstruction process. Furthermore, it often leads to simple and direct solutions for the estimation of both camera and scene parameters, which may eventually be adjusted non-linearly for highest accuracy. In addition, it enables modeling from small sets of images, in particular from single images, thus making possible reconstructions from images not originally taken for that purpose, like archival images or images from the internet for instance.

The literature on using prior information for self-calibration and for metric reconstruction is vast and widely scattered. An exhaustive review of all existing approaches is therefore beyond the scope of this article, and we will concentrate on works which have somehow inspired the method we propose, especially the direct approaches giving a good first estimate of the camera and scene parameters. Among many possible criteria for classifying modeling methods in this context, two appear to be more natural: first, the type of information which is used and second, how this information is used. There is a large variety of information which can be incorporated into a 3D modeling process. It can be simple knowledge on camera intrinsic parameters or poses (stationarity, translation, etc.) or on 3D scene structures (calibration patterns); it can also be the complete knowledge of calibration primitives and scene elements such as points, lines and planes, as well as a partial knowledge of high-level primitives like cubes, prisms, cylinders, etc. Nonetheless, note that whatever the information is, it can often be used at any stage of the 3D modeling process, including the initial intrinsic calibration, the pose estimation, the model reconstruction or even an additional non-linear adjustment of the initial estimate at every step.

The first class of approaches we mention is based on known positions of points in the 3D space, or known calibration patterns [50]. Unfortunately, such information relies on specific acquisition systems and is therefore seldom available in general situations. The use of prior knowledge on some intrinsic camera parameters, i.e. self-calibration, offers the opportunity to build much more flexible systems. In standard self-calibration algorithms [30, 48, 18, 33], the 3D reconstruction is done in 3 steps, recovering, in order, the projective, affine and Euclidean strata, the projective–affine step being considered as the most non-linear and thus most difficult step. One of the main problems are the critical motion sequences, for which the self-calibration does not have a unique solution [40]. This problem has been dealt with by restraining the camera motions [17, 10, 2] or by incorporating prior knowledge on the camera [58] or the scene into the calibration process. To get stable results for self-calibration, an important number of images is usually necessary.

A large variety of geometric constraints can disambiguate projective reconstructions from metric ones and allow to decrease the number of images required to obtain a satisfying reconstruction. Many of them can easily be incorporated into a self-calibration framework. A very common constraint in this category is given by vanishing points of mutually orthogonal directions, as defined by known cuboidal structures [7, 9, 8] or by dominating scene directions [24]. Also, knowing the metric structure of scene planes appears to be useful in this context, through rectified planes [26], maps [5] or known 3D plane–image homographies [42, 57]. It is also possible to use multiple images of unknown planes, however more images in general positions are needed in that case [49, 27].

Another category of approaches allows for the simultaneous recovery of cameras and 3D models. Factorization methods have been successfully applied to points [46, 44, 28, 39], lines [47, 29] and planes [38, 41]. Even though accurate, factorization algorithms suffer from missing data, i.e. when one primitive is not present in all the concerned images. That is why various systems based on image tensors have been proposed [3, 14]. These systems reconstruct the scene structure using 2, 3 or 4 images, and eventually sequentially or hierarchically join these reconstructed parts into consistent models. An overview of this can be found in [20].

When cameras are calibrated, it is relatively easy to reconstruct 3D structure. However, and as mentioned previously, using geometric constraints improves dramatically the reconstruction quality, especially when a single or only few images are considered. Even simple constraints can be very efficient for this purpose. In [1, 43], vanishing lines of planes and coplanarity constraints are used for single image reconstruction. More complex systems, dealing with multiple images, use constraints such as coincidence, parallelism, orthogonality, etc., between points, lines, and planes [35, 21, 16, 53, 55].

Most of the mentioned methods give solutions for both reconstruction and calibration. However, it is usually advantageous to optimize the obtained parameters using non-linear methods in an additional step. One way to incorporate geometrical information in this process is to augment the cost function by the penalty terms corresponding to the model constraints [31] or use the constrained optimization techniques [45, 32, 51]. The first type of methods however do not always ensure that the final structure will respect the given constraints exactly, and the constrained optimization techniques are in general designed for small equation systems, sometimes not sufficient to describe all the 3D scene dependencies. In answer to this, some systems rather try to find a minimal subset of parameters such that the associated scene models conform to the geometric constraints, and then, optimize over that subset. This includes methods which use simple primitives and constraints to compute the minimal scene parameterization [4, 55], as well as methods which are based on high-level scene descriptions through complex primitives like cubes, prisms, cylinders, etc. [11]. Recently, some effort has been devoted to the automatic detection of such primitives [12, 52]. The methods in this category ensure, by the strong inherent geometric constraints, that the final models are visually correct. On the other hand, they require scenes to be composed of basic primitives only, which is not always the case, see for example most archaeological environments.

In this paper, we address the first part of the 3D model acquisition process, the intrinsic and extrinsic camera calibration. In particular, we study the use of a specific calibration primitive: the parallelepipeds. Parallelepipeds are frequently present in man-made environments and they naturally encode the affine structure of the scene. Any information about their Euclidean structure (angles or ratios of edge lengths), possibly combined with information about camera parameters allows for Euclidean reconstruction. We propose an elegant formalism to incorporate such information, in which camera parameters are dual to parallelepiped parameters, i.e. any knowledge about one entity provides constraints on the parameters of the others. Consequently, the image of a known parallelepiped defines the camera parameters, and reciprocally, a calibrated image of a parallelepiped defines its Euclidean shape (up to its size). In this paper, we synthesize our work on parallelepipeds [53, 54] and propose more elegant and efficient approaches.

The cameras and parallelepipeds parameters are recovered in two steps. Firstly, the factorization based approach is used to compute their intrinsic and orientation parameters. Interestingly, the use of the three-dimensional structure allows to deal very easily with the classical problems of the factorization methods: the unknown scale factors and the missing data. Then the least square optimization is used to recover simultaneously the scale and position parameters in the common euclidean frame. The use of the well-constrained calibration primitives allows to obtain good calibration results even from very small number of images, however, depending on the information about the scene, the singularities might occur. They are collected into a detailed dictionary, accompanied by the sketch of the methodology used for its construction.

Our calibration approach is conceptually close to self-calibration methods, especially to methods for upgrading affine to Euclidean structure [18, 33] or methods that consider special camera motions [17, 10, 2]. The way the metric information of a parallelepiped is used, is also similar to the vanishing point based methods [7, 9, 8, 24]. Some properties of our algorithm are also common with plane-based approaches [42, 57, 49, 27, 38, 41]. While more flexible than standard calibration techniques, homography-based approaches still require either Euclidean information or, for self-calibration, many images in general position [48], or at least one plane visible in all images [37]. In this sense, our approach is a generalization of plane-based methods with metric information, to three-dimensional parallelepipedic patterns. This allows to handle missing data and unknown scale factors and simplifies the formulation of calibration constraints.

While the main contributions of the paper concern the estimation of camera and parallelepiped parameters, we also propose a method for enhancing reconstructions with primitives other than parallelepipeds. The complete system allows for both calibration and 3D model acquisition from a small number of arbitrary images, and this with a reasonable amount of user interaction.

The paper is organized as follows. Section 2 gives definitions and some background. Section 3 introduces the concept of camera–parallelepiped duality. Calibration using parallelepipeds and a study on the singular configurations are described in sections 4 and 5.

Sections 6 and 7 describe our approaches for pose estimation and 3D reconstruction. Experimental results are then presented in section 8 before concluding.

## 2 Preliminaries

### 2.1 Camera parameterization

We represent cameras using the pinhole model. The projection from a 3D point  $\mathbf{P}$  to its 2D image point  $\mathbf{p}$  is expressed by:  $\mathbf{p} \sim \mathbf{M}\mathbf{P}$ , where  $\mathbf{M}$  is a  $3 \times 4$  matrix, which can be decomposed as:

$$\mathbf{M} = \mathbf{K}(\mathbf{R} \quad \mathbf{t})$$

The  $3 \times 4$  matrix  $(\mathbf{R} \quad \mathbf{t})$  encapsulates the camera's pose in the world coordinate system, or its extrinsic parameters: the rotation matrix  $\mathbf{R}$  represents its orientation and the vector  $\mathbf{t}$  its position. The matrix  $\mathbf{K}$  is the  $3 \times 3$  calibration matrix containing the camera's intrinsic parameters:

$$\mathbf{K} = \begin{pmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (1)$$

where  $\alpha_u$  and  $\alpha_v$  stand for the focal length, expressed in horizontal and vertical pixel dimensions,  $s$  is a skew parameter considered here as equal to zero, and  $(u_0, v_0)$  are the pixel coordinates of the principal point. The notation  $\tau = \frac{\alpha_u}{\alpha_v}$  will be used for the camera aspect ratio. The term camera axes will be used for axes of the coordinate system attached to the camera projection center, where two axes are parallel to pixel edges and the third one is orthogonal to the image plane (the viewing axis). In the following, we will also use the IAC (image of the absolute conic) representation of the intrinsic parameters, namely the matrix  $\omega \sim \mathbf{K}^{-\top}\mathbf{K}^{-1}$ :

$$\omega \sim \begin{pmatrix} 1 & 0 & -u_0 \\ 0 & \tau^2 & -\tau^2 v_0 \\ -u_0 & -\tau^2 v_0 & \tau^2 \alpha_v^2 + u_0^2 + \tau^2 v_0^2 \end{pmatrix}. \quad (2)$$

### 2.2 Parallelepiped parameterization

A parallelepiped is defined by twelve parameters: six extrinsic parameters describing its orientation and position, and six intrinsic parameters describing its Euclidean shape: three dimension parameters (edge lengths  $l_1, l_2$  and  $l_3$ ) and three angles between edges ( $\theta_{12}, \theta_{23}, \theta_{13}$ ). These intrinsic parameters are illustrated in figure 1. The parallelepiped may be represented compactly in matrix form by a  $4 \times 4$  matrix  $\mathbf{N}$ :

$$\mathbf{N} = \begin{pmatrix} \mathbf{S} & \mathbf{v} \\ \mathbf{0}^\top & 1 \end{pmatrix} \tilde{\mathbf{L}}$$

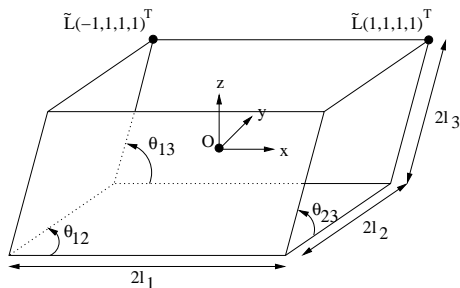


Figure 1: Parameterization of a parallelepiped:  $2l_i$  are the edge lengths;  $\theta_{ij}$  are the angles between non-parallel edges.

where  $S$  is a rotation matrix and  $\mathbf{v}$  a vector, representing the parallelepiped's pose (extrinsic parameters). The  $4 \times 4$  matrix  $\tilde{L}$  represents the parallelepiped's shape:

$$\tilde{L} = \begin{pmatrix} l_1 & l_2 c_{12} & l_3 c_{13} & 0 \\ 0 & l_2 s_{12} & l_3 \frac{c_{23} - c_{13} c_{12}}{s_{12}} & 0 \\ 0 & 0 & l_3 \sqrt{\frac{s_{12}^2 - c_{13}^2 s_{12}^2 - (c_{23} - c_{13} c_{12})^2}{s_{12}^2}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with:  $c_{ij} = \cos \theta_{ij}$ ,  $s_{ij} = \sin \theta_{ij}$ ,  $\theta_{ij} \in ]0, \pi[$ ,  $l_i > 0$ .

The matrix  $\tilde{L}$  represents the affine transformation between a canonic cube and a parallelepiped with the given shape. Concretely, a vertex  $(\pm 1, \pm 1, \pm 1, 1)^T$  of the canonic cube is mapped, by  $\tilde{L}$ , to a vertex of our parallelepiped's intrinsic shape. Then, the pose part of  $N$  maps the vertices into the world coordinate system.

Other parameterizations for  $\tilde{L}$  may be chosen, but the above one is attractive due to its upper triangular form. This underlines the fact that  $\tilde{L}$  plays the same role for the parallelepiped as the calibration matrix  $K$  for a camera.

The analogous entity to a camera's IAC  $\omega$ , is the matrix  $\mu$ , defined by:

$$\mu \sim L^T L \sim \begin{pmatrix} l_1^2 & l_1 l_2 \cos \theta_{12} & l_1 l_3 \cos \theta_{13} \\ l_1 l_2 \cos \theta_{12} & l_2^2 & l_2 l_3 \cos \theta_{23} \\ l_1 l_3 \cos \theta_{13} & l_2 l_3 \cos \theta_{23} & l_3^2 \end{pmatrix}, \quad (3)$$

where  $L$  is the upper left  $3 \times 3$  matrix of  $\tilde{L}$ .

Hence, there is a seemingly perfect symmetry between intrinsic parameters of cameras and parallelepipeds. The only difference is that in some cases, the *size* of a parallelepiped matters, as will be explained in the following. As for cameras, the fact that  $K_{33} = 1$  allows to fix the scale factor in the relation  $\omega \sim K^{-T} K^{-1}$ , and thus to extract  $K$  uniquely from the IAC  $\omega$ , e.g. using Cholesky decomposition. As for parallelepipeds however, we have no such constraint on its "calibration matrix"  $L$ , so the relation  $\mu \sim L^T L$  gives us a parallelepiped's

Euclidean shape, but not its (absolute) size. This does not matter in general, since we are usually only interested in reconstructing a scene up to some scale. However, when reconstructing several parallelepipeds, one needs to recover at least their *relative* sizes.

There are many possibilities to define the size of parallelepipeds. We choose the following definition, due to its appropriateness in the equations underlying our calibration and reconstruction algorithms below: the **size** of a parallelepiped is defined as

$$s = (\det L)^{1/3} .$$

This definition is actually directly linked to the parallelepiped's volume:  $s^3 = \det L = \text{Vol}/8$  (the factor 8 arises since our canonic cube has an edge length of 2).

### 3 Projections of Parallelepipeds

#### 3.1 One Parallelepiped in A Single View

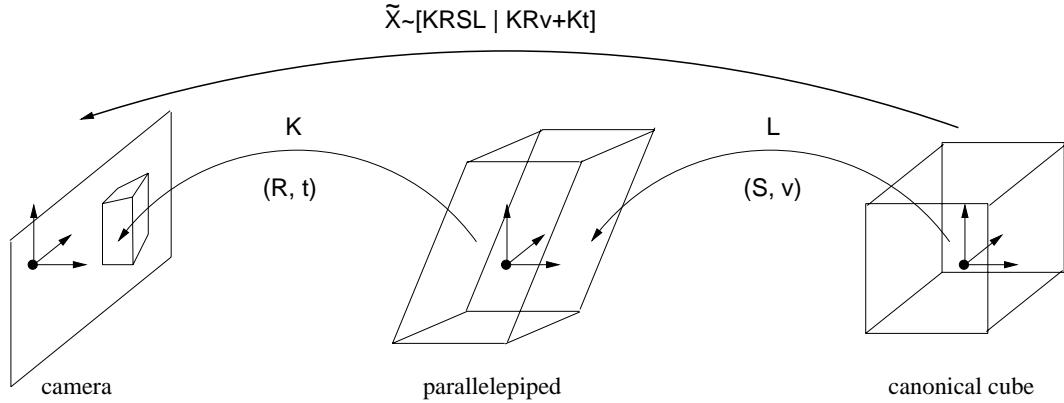


Figure 2: The projection of the canonic parallelepiped (cube) into the image. Matrices  $K$ ,  $L$  correspond to intrinsic parameters of camera and parallelepiped and  $(R, t)$ ,  $(S, v)$  correspond to extrinsic parameters of camera and parallelepiped, respectively.

In this section, we introduce the concept of duality between the intrinsic characteristics of a camera and those of a parallelepiped.

To do so, consider the projection of the parallelepiped's vertices into the camera. Let  $C_i, i = 1..8$  be the homogeneous coordinates of the canonic cube's vertices. As described in section 2.2, the corresponding vertex of the parallelepiped is given as:

$$P_i = NC_i = \begin{pmatrix} S & v \\ \mathbf{0}^T & 1 \end{pmatrix} \tilde{L}C_i$$

and its image point is:

$$p_i \sim MP_i = K(R \ t) \begin{pmatrix} S & v \\ \mathbf{0}^T & 1 \end{pmatrix} \tilde{L}C_i . \quad (4)$$

In the above equation, we define the *canonic projection matrix*:

$$\tilde{X} \sim K(R \ t) \begin{pmatrix} S & v \\ \mathbf{0}^T & 1 \end{pmatrix} \tilde{L} . \quad (5)$$

This matrix represents a perspective projection that maps the vertices of the canonic cube onto the image points of the parallelepiped's vertices. This is illustrated on the figure 2. Given image points for sufficiently many vertices<sup>1</sup>, the canonic projection matrix

<sup>1</sup>Five image points and one image direction are in general sufficient. Additional points make the computation more stable.

can be computed, even in the absence of prior knowledge on intrinsic or extrinsic parameters. Our calibration and pose estimation algorithms are based on the link between the canonic projection matrix (which we suppose given from now on) and the camera's and parallelepiped's intrinsic and extrinsic parameters.

Let us consider this in more detail. First, we may identify the *relative pose* between camera and parallelepiped in (5), represented by the following  $3 \times 4$  matrix:

$$(\mathbf{R} \quad \mathbf{t}) \begin{pmatrix} \mathbf{S} & \mathbf{v} \\ \mathbf{0}^\top & 1 \end{pmatrix} = (\mathbf{RS} \quad \mathbf{Rv} + \mathbf{t})$$

Second, let us consider the leading  $3 \times 3$  sub-matrix  $\mathbf{X}$  of the canonic projection matrix  $\tilde{\mathbf{X}}$ , which is given by:

$$\mathbf{X} \sim \mathbf{K}(\mathbf{RS})\mathbf{L}. \quad (6)$$

Due to the orthogonality of the rotation matrices  $\mathbf{R}$  and  $\mathbf{S}$ , it is simple to derive the following relation between the camera's IAC  $\omega$  and the corresponding entity  $\mu$  of the parallelepiped:

$$\mathbf{X}^\top \omega \mathbf{X} \sim \mu. \quad (7)$$

This equation establishes an interesting duality between the intrinsic parameters of a camera and those of a parallelepiped. It shows (unsurprisingly) that knowing the parallelepiped's shape  $\mu$  allows to calibrate the camera. Conversely, knowing the camera's intrinsic parameters allows to directly compute the parallelepiped's Euclidian shape, also from a single image.

In the next sections, we generalize the use of this duality for calibration and pose estimation to the case of multiple parallelepipeds seen in multiple cameras and to the use of *partial* knowledge about the camera's or parallelepiped's intrinsic parameters. Before doing so, let us describe a few interesting links between our and other (self-) calibration scenarios.

Classical self-calibration proceeds usually in two main steps: first, a projective 3D reconstruction of the scene is obtained from correspondences across two or more images. Then, the projective reconstruction is transformed to a Euclidean one using the available prior knowledge on intrinsic parameters. This upgrade is sometimes interlaced by an intermediate upgrade to an affine reconstruction.

In our scenario, we have a 3D reconstruction of the scene already from a *single* rather than multiple images, which is furthermore of *affine* rather than projective nature: we know that the observed parallelepiped's shape is that of a cube, up to some affine transformation. Analogously, our canonic projection matrix is equal to the true one up to an affine transformation. Hence, self-calibration in our scenario does not need to recover the plane at infinity, which is known to be the hardest ("most non-linear") part of classical self-calibration. Indeed, our calibration method is somewhat similar to the affine-to-Euclidean upgrade of stratified self-calibration approaches, e.g. [18, 33].

Similarities also exist with (self-) calibration approaches based on special camera motions: calibrating a rotating camera [17, 10] is more or less equivalent to self-calibrating a camera in general motion once affine structure is known. Other approaches recover the affine structure

by first performing pure translations and then general motions or to approaches that consider special camera motions [2, 34].

Our scenario is similar to these. In the following sections we show how it allows to efficiently combine the usual self-calibration constraints with constraints on scene structure. This enables to perform calibration (and 3D reconstruction) from very few images; one image may actually be sufficient.

### 3.2 $n$ Parallelepipeds in $m$ Views

The main motivation for the work described in this paper is to generalize the use of the duality introduced in the previous section: we consider the general case where multiple parallelepipeds are seen by multiple cameras (not all parallelepipeds need to be seen by all cameras). Furthermore, we do not in general suppose that some cameras or parallelepipeds are fully calibrated. We rather want to make efficient and complete use of any kind of partial calibration information. As for the cameras, this amounts to partial knowledge on their intrinsic parameters that is routinely used for self-calibration. As for parallelepipeds, we rather consider them as “vehicles” to jointly express simple yet useful geometric scene constraints. Defining orthogonality or parallelism constraints between for example “only” pairs of lines, amounts to providing information about the structure of individual *planes* in the scene. Parallelepipeds however allow to directly express richer couplings of constraints on *3D* scene structure.

Let us now consider the general case where  $n$  parallelepipeds are seen by  $m$  cameras. Let  $\tilde{X}_{ik}$  be the canonic projection matrix associated with the projection of the  $k$ th parallelepiped in the  $i$ th camera:

$$\tilde{X}_{ik} \sim K_i (R_i \quad \mathbf{t}_i) \begin{pmatrix} S_k & \mathbf{v}_k \\ \mathbf{0}^\top & 1 \end{pmatrix} \tilde{L}_k$$

Let us explicitly introduce scale factors  $\lambda_{ik}$  such that the equality up to scale in the above equation can be turned into a component-wise equality:

$$\lambda_{ik} \tilde{X}_{ik} = K_i (R_i \quad \mathbf{t}_i) \begin{pmatrix} S_k & \mathbf{v}_k \\ \mathbf{0}^\top & 1 \end{pmatrix} \tilde{L}_k \quad (8)$$

We may group together these equations for all  $m$  cameras and  $n$  parallelepipeds, into the following single matrix equation:

$$\begin{aligned}
& \underbrace{\begin{bmatrix} \lambda_{11}\tilde{X}_{11} & \cdots & \lambda_{1n}\tilde{X}_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{m1}\tilde{X}_{m1} & \cdots & \lambda_{mn}\tilde{X}_{mn} \end{bmatrix}}_{\mathcal{X}_{3m \times 4n}} = \\
& \underbrace{\begin{bmatrix} K_1 (R_1 & \mathbf{t}_1) \\ \vdots \\ K_m (R_m & \mathbf{t}_m) \end{bmatrix}}_{\mathcal{M}_{3m \times 4}} \underbrace{\left[ \left( \begin{smallmatrix} S_1 & \mathbf{v}_1 \\ \mathbf{0}^\top & 1 \end{smallmatrix} \right) \tilde{L}_1 \cdots \left( \begin{smallmatrix} S_n & \mathbf{v}_n \\ \mathbf{0}^\top & 1 \end{smallmatrix} \right) \tilde{L}_n \right]}_{\mathcal{S}_{4 \times 4n}} \quad (9)
\end{aligned}$$

This equation naturally leads to the idea of a factorization-based calibration algorithm, which will be developed in section 4. It is based on the following observation. The matrix  $\mathcal{X}$  contains all information that can be recovered from the parallelepipeds' image points alone (below, we discuss the issue of computing the scale factors  $\lambda_{ik}$ ). In analogy with [46], we call it the *measurement matrix*. Since the measurement matrix is the product of a “motion matrix”  $\mathcal{M}$  of 4 columns, with a “shape matrix”  $\mathcal{S}$  of 4 rows, its rank can be 4 at most (in the absence of noise).

We might aim at extracting intrinsic and extrinsic parameters of cameras and parallelepipeds directly from a rank-4-factorization of  $\mathcal{X}$ . One step of many factorization methods for structure and motion recovery is to disambiguate the result of the factorization: in general, for a rank- $r$ -factorization, motion and shape are recovered up to a transformation represented by an  $r \times r$  matrix (in our case, this would be a 3D projective transformation). The ambiguity can be reduced using e.g. constraints on intrinsic camera parameters (see more details in section 4). In our case, we observe that the  $4 \times 4$  sub-blocks of the shape matrix  $\mathcal{S}$  are affine transformations (last row consists of three zeroes and a one). We would have to include this constraint into the disambiguation, but nevertheless, the result would not in general exactly satisfy the affine form of these sub-blocks. We thus cut the problem in two steps, which allows to easily guarantee that the sub-blocks of the shape matrix be affine transformations. In the first step (section 4), we consider a “reduced measurement matrix” consisting of the leading  $3 \times 3$  sub-matrices of the  $\tilde{X}_{ik}$ . We extract *intrinsic parameters* and *orientation* of our cameras and parallelepipeds based on a rank-3-factorization and a disambiguation stage using calibration and scene constraints. In the second step (section 6), we then estimate the *position* of cameras and parallelepipeds, as well as the parallelepipeds' *size*.

Just as a sidenote, we observe that, for two views  $i$  and  $j$ , and a parallelepiped  $k$ , the infinite homography between the two views is given by the product  $X_{ik}X_{jk}^{-1}$ .

## 4 Estimating Intrinsic and Orientation Parameters by Factorization

In this section we concentrate on the computation of the cameras' and parallelepipeds' intrinsic parameters and orientation (rotation), based on equation (9) and the observations concerning it, cf. the previous section. As mentioned previously, we first restrict our attention to the leading  $3 \times 3$  submatrices of the  $\tilde{\mathbf{X}}_{ik}$ , like we did in section 3.1 for the establishment of the duality between intrinsic parameters of cameras and parallelepipeds. We thus deal with the corresponding subpart of equation (9):

$$\underbrace{\begin{bmatrix} \lambda_{11}\mathbf{X}_{11} & \cdots & \lambda_{1n}\mathbf{X}_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{m1}\mathbf{X}_{m1} & \cdots & \lambda_{mn}\mathbf{X}_{mn} \end{bmatrix}}_{\mathcal{X}'_{3m \times 3n}} = \underbrace{\begin{bmatrix} \mathbf{K}_1\mathbf{R}_1 \\ \vdots \\ \mathbf{K}_m\mathbf{R}_m \end{bmatrix}}_{\mathcal{M}'_{3m \times 3}} \underbrace{\begin{bmatrix} \mathbf{S}_1\mathbf{L}_1 & \cdots & \mathbf{S}_n\mathbf{L}_n \end{bmatrix}}_{\mathcal{S}'_{3 \times 3n}} \quad (10)$$

In the following sections, we describe the different steps of our factorization-based method. We first deal with the important problem of missing data. Then it is described how the scale factors  $\lambda_{ik}$  needed to construct the measurement matrix  $\mathcal{X}'$ , are computed. The factorization itself is described in section 4.3, followed by the most important aspect: how to disambiguate the factorization's result in order to extract intrinsic and orientation parameters. A summary of these steps and a discussion of minimal cases and singularities is provided at the end of this section. The computation of position parameters and parallelepiped size is dealt with in section 6.

### 4.1 Missing Data

As with all factorization problems, our method might suffer from the problem of missing data, i.e. missing  $\mathbf{X}_{ik}$ . Indeed, in practice, the condition that all parallelepipeds are seen in all views can not always be satisfied. However, each missing matrix  $\mathbf{X}_{ik}$  can be deduced from others if there is one camera  $j$  and one parallelepiped  $l$  such that the transformations  $\mathbf{X}_{jl}$ ,  $\mathbf{X}_{jk}$  and  $\mathbf{X}_{il}$  are known. The missing matrix can then be computed using:

$$\mathbf{X}_{ik} \sim \mathbf{X}_{il} (\mathbf{X}_{jl})^{-1} \mathbf{X}_{jk}. \quad (11)$$

Several equations of this type may be used simultaneously to increase the accuracy. In that case, care has to be taken since equation (11) is defined up to scale only. This problem can be circumvented very simply though, by normalizing all  $\mathbf{X}_{ik}$  to unit determinant.

These observations motivate a simple recursive method<sup>2</sup> to compute missing matrices  $\mathbf{X}_{ik}$ : at each iteration, we compute the one for which most equations of type (11) are available. Previously computed matrices  $\mathbf{X}_{ik}$  can be involved at every successive iteration of this procedure.

<sup>2</sup>Compare with the analogous method in [41].

## 4.2 Recovery of Scale Factors

The reduced measurement matrix  $\mathcal{X}'$  in (10) is, in the absence of noise, of rank 3, being the product of a matrix of 3 columns and a matrix of 3 rows. This however only holds if a correct set of scale factors  $\lambda_{ik}$  is used. For other problems, these are often non trivial to compute, see e.g. [27, 44, 56]. In our case however, this turns out to be rather simple.

Let us first write  $\mathbf{A}_i = \mathbf{K}_i \mathbf{R}_i$  and  $\mathbf{B}_k = \mathbf{S}_k \mathbf{L}_k$ . What we know is that (in the absence of noise), there exist matrices  $\mathbf{A}_i, i = 1..m$  and  $\mathbf{B}_k, k = 1..n$  such that:

$$\forall i, k : \mathbf{X}_{ik} \sim \mathbf{A}_i \mathbf{B}_k$$

Since this equation is valid up to scale only, we also have:

$$\forall i, k : \mathbf{X}_{ik} \sim (a_i \mathbf{A}_i) (b_k \mathbf{B}_k)$$

for any non-zero scale factors  $a_i, i = 1..m$  and  $b_k, k = 1..n$ . Consequently, this is also true for the scale factors with:

$$\begin{aligned} \det(a_i \mathbf{A}_i) &= 1 \\ \det(b_k \mathbf{B}_k) &= 1 \end{aligned}$$

Note that we do not need to know these scale factors; it is sufficient to know they exist. Hence:

$$\forall i, k : \mathbf{X}_{ik} \sim a_i b_k \mathbf{A}_i \mathbf{B}_k,$$

with  $\det(a_i b_k \mathbf{A}_i \mathbf{B}_k) = \det(a_i \mathbf{A}_i) \det(b_k \mathbf{B}_k) = 1$ .

To achieve a component-wise equality  $\lambda_{ik} \mathbf{X}_{ik} = (a_i \mathbf{A}_i) (b_k \mathbf{B}_k)$ , we need to use scale factors<sup>3</sup>  $\lambda_{ik}$  such that  $\det(\lambda_{ik} \mathbf{X}_{ik}) = 1$ . Hence:

$$\lambda_{ik} = (\det \mathbf{X}_{ik})^{-1/3}$$

In the following, we assume that the  $\mathbf{X}_{ik}$  are already normalized to unit determinant, i.e. that  $\lambda_{ik} = 1$ . Equation (10) becomes:

$$\underbrace{\begin{bmatrix} \mathbf{X}_{11} & \cdots & \mathbf{X}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{X}_{m1} & \cdots & \mathbf{X}_{mn} \end{bmatrix}}_{\mathcal{X}'_{3m \times 3n}} = \underbrace{\begin{bmatrix} a_1 \mathbf{K}_1 \mathbf{R}_1 \\ \vdots \\ a_m \mathbf{K}_m \mathbf{R}_m \end{bmatrix}}_{\mathcal{M}'_{3m \times 3}} \underbrace{\begin{bmatrix} b_1 \mathbf{S}_1 \mathbf{L}_1 & \cdots & b_n \mathbf{S}_n \mathbf{L}_n \end{bmatrix}}_{\mathcal{S}'_{3 \times 3n}} \quad (12)$$

The scale factors  $a_i$  and  $b_k$  do not matter for now; all that counts is that the measurement matrix  $\mathcal{X}'$  containing the normalized  $\mathbf{X}_{ik}$ , is of rank 3 at most, and can thus be factorized as shown below.

<sup>3</sup>It is well known that two non-singular  $3 \times 3$  matrices that are equal up to scale and whose determinants are equal, are also equal component-wise.

### 4.3 Factorization

As usual, we use the SVD (Singular Value Decomposition) to obtain the low-rank factorization of the measurement matrix. Let the SVD of  $\mathcal{X}'$  be given as:

$$\mathcal{X}'_{3m \times 3n} = \mathbf{U}_{3m \times 3n} \mathbf{\Sigma}_{3n \times 3n} \mathbf{V}_{3n \times 3n}^\top$$

The diagonal matrix  $\mathbf{\Sigma}$  contains the singular values of  $\mathcal{X}'$ . Let them be ordered:  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{3n}$ . In the absence of noise,  $\mathcal{X}'$  is of rank 3 at most and  $\sigma_4 = \dots = \sigma_{3n} = 0$ . If noise is present,  $\mathcal{X}'$  is of full rank in general. Setting all singular values to zero, besides the three largest ones, leads to the best rank-3 approximation of  $\mathcal{X}'$  (in the sense of the Frobenius norm [36]).

In the following, we consider the decomposition of the rank-3 approximation of  $\mathcal{X}'$  (for ease of notation, we denote this also as  $\mathcal{X}'$ ):

$$\mathcal{X}' = \mathbf{U}_{3m \times 3n} \text{diag}(\sigma_1, \sigma_2, \sigma_3, 0, \dots, 0) \mathbf{V}_{3n \times 3n}^\top$$

In the matrix product on the right, only columns of  $\mathbf{U}$  and rows of  $\mathbf{V}^\top$  that correspond to non-zero singular values, contribute. Hence:

$$\mathcal{X}' = \mathbf{U}'_{3m \times 3} \begin{pmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{pmatrix} (\mathbf{V}'^\top)_{3 \times 3n}$$

where  $\mathbf{U}'$  (resp.  $\mathbf{V}'$ ) consists of the first three columns of  $\mathbf{U}$  (resp.  $\mathbf{V}$ ). Let us define

$$\begin{aligned} \mathbf{U}'' &= \mathbf{U}' \begin{pmatrix} \sqrt{\sigma_1} & & \\ & \sqrt{\sigma_2} & \\ & & \sqrt{\sigma_3} \end{pmatrix} \\ \mathbf{V}'' &= \mathbf{V}' \begin{pmatrix} \sqrt{\sigma_1} & & \\ & \sqrt{\sigma_2} & \\ & & \sqrt{\sigma_3} \end{pmatrix} \end{aligned}$$

Thus:

$$\mathcal{X}' = \mathbf{U}'' \mathbf{V}''^\top$$

This represents a decomposition of the measurement matrix  $\mathcal{X}'$  into a product of a matrix of 3 columns with a matrix of 3 rows. Note however, that this decomposition is not unique. For any non-singular  $3 \times 3$  matrix  $\mathbf{T}$ , the following is also a valid decomposition:

$$\mathcal{X}' = (\mathbf{U}'' \mathbf{T}^{-1}) (\mathbf{T} \mathbf{V}''^\top)$$

Making the link with equation (12), we obtain:

$$\begin{bmatrix} a_1 \mathbf{K}_1 \mathbf{R}_1 \\ \vdots \\ a_m \mathbf{K}_m \mathbf{R}_m \end{bmatrix} [b_1 \mathbf{S}_1 \mathbf{L}_1 \quad \dots \quad b_n \mathbf{S}_n \mathbf{L}_n] = (\mathbf{U}'' \mathbf{T}^{-1}) (\mathbf{T} \mathbf{V}''^\top) \quad (13)$$

Let us write  $U''$  and  $V''$  as follows as a composition of  $3 \times 3$  submatrices:

$$U'' = \begin{bmatrix} U_1 \\ \vdots \\ U_m \end{bmatrix} \quad V'' = \begin{bmatrix} V_1 \\ \vdots \\ V_n \end{bmatrix}$$

Equation (13) thus becomes:

$$\begin{bmatrix} a_1 K_1 R_1 \\ \vdots \\ a_m K_m R_m \end{bmatrix} [b_1 S_1 L_1 \cdots b_n S_n L_n] = \begin{bmatrix} U_1 T^{-1} \\ \vdots \\ U_m T^{-1} \end{bmatrix} [TV_1^T \cdots TV_n^T] \quad (14)$$

How to estimate  $T$  is explained in section 4.4. Once a correct estimate is given, we can directly extract the matrices  $A_i = a_i K_i R_i$  and  $B_k = b_k S_k L_k$ , from which in turn the individual rotation matrices and calibration matrices can be recovered using a Cholesky or QR-decomposition. The Cholesky decomposition of  $A_i A_i^T$  for example, results in an upper triangular matrix  $M_i = a_i K_i$ . Based on the requirement  $K_{i,33} = 1$ , we can compute the unknown scale factor  $a_i$  as  $a_i = M_{i,33}$ . The calibration matrix is finally obtained as  $K_i = \frac{1}{a_i} M_i$ .

As for the parallelepipeds, we do not have any constraint on the entries of their calibration matrices  $L_k$ . Hence, we can compute them only up to the unknown scale factors  $b_k$ . This means that we can compute the *shape* of each parallelepiped, but not (yet) their *size* (or, volume). In section 6, we explain how to compute their (relative) size.

We now briefly discuss the structure and geometric signification of the matrix  $T$ . Note that  $T$  actually represents the non-translational part of a 3D affine transformation (its upper left  $3 \times 3$  submatrix). This is just another expression of the previously mentioned fact that due to the observation of parallelepipeds, we directly have an affine reconstruction (of scene and cameras).

The matrix  $T$  can only be computed up to an arbitrary rotation and scale: for any rotation matrix  $R$  and scale factor  $s$ ,  $T' = sRT$  can not be distinguished from  $T$  in the factorization since  $T'^{-1}T' \sim T^{-1}T$ . This ambiguity is natural and expresses the fact that the global Euclidean reference frame for the reconstruction of our parallelepipeds and cameras can be chosen arbitrarily. Without loss of generality, we may thus assume that  $T$  is upper triangular. This highlights the fact that our estimation problem has only 5 degrees of freedom (6 parameters for an upper triangular  $3 \times 3$  matrix minus one for the arbitrary scale) which can also be explained in more geometric terms: as explained previously, our problem is somewhat equivalent to self-calibration with known affine structure. The 5 degrees of the problem may be interpreted as the coefficients of the absolute conic on the plane at infinity.

#### 4.4 Disambiguating the Factorization

We now deal with the estimation of the unknown transformation  $T$  appearing in equation (14). As will be seen below, and as is often the case in self-calibration problems, it is

simpler to not directly estimate  $\mathbb{T}$ , but the symmetric and positive definite  $3 \times 3$  matrix  $\mathbb{Z}$  defined as:

$$\mathbb{Z} = \mathbb{T}^\top \mathbb{T} \quad (15)$$

We may observe that this matrix represents the absolute conic on the plane at infinity. Once  $\mathbb{Z}$  is estimated,  $\mathbb{T}$  may be extracted from it using Cholesky decomposition. As described above,  $\mathbb{T}$  is defined up to a rotation and scale, so the upper triangular Cholesky factor of  $\mathbb{Z}$  can directly be used as the estimate for  $\mathbb{T}$ .

The matrix  $\mathbb{Z}$  (and thus  $\mathbb{T}$ ), can be estimated in various ways, using any information about the cameras or the parallelepipeds, e.g. prior knowledge on relative positioning of some entities. Here, we concentrate on exploiting prior information on intrinsic parameters, of both, cameras and parallelepipeds. There are two types of information that we consider:

- knowledge of the actual value of some intrinsic parameter for some camera or parallelepiped.
- knowledge that two or more cameras (or parallelepipeds) have the same value for some intrinsic parameter. We also sometimes speak of “constant” intrinsic parameters.

We first describe the use of prior knowledge on camera intrinsics.

#### 4.4.1 Using Knowledge on Camera Intrinsics

From equation (14), we have:

$$a_i \mathbf{K}_i \mathbf{R}_i = \mathbf{U}_i \mathbb{T}^{-1}$$

Due to the orthogonality of  $\mathbf{R}_i$ , we get:

$$a_i^2 \underbrace{\mathbf{K}_i \mathbf{K}_i^\top}_{\omega_i^{-1}} = \mathbf{U}_i \underbrace{\mathbb{T}^{-1} \mathbb{T}^{-\top}}_{\mathbb{Z}^{-1}} \mathbf{U}_i^\top$$

Neglecting the unknown scale factor  $a_i$  and taking the inverse of both sides of the equation, we obtain (note that the  $\mathbf{U}_i$  are not orthogonal in general):

$$\omega_i \sim \mathbf{U}_i^{-\top} \mathbb{Z} \mathbf{U}_i^{-1}. \quad (16)$$

We are now ready to formulate constraints on  $\mathbb{Z}$  based on prior knowledge on the cameras' intrinsics.

**Known values of camera intrinsics** Knowing respectively the aspect ratio and principal point coordinates of a camera  $i$  gives the following constraints on its IAC  $\omega_i$  (based on equation (2)):

$$\begin{aligned} \tau_i^2 \omega_{i,11} - \omega_{i,22} &= 0 \\ u_{i,0} \omega_{i,11} + \omega_{i,13} &= 0 \\ v_{i,0} \omega_{i,22} + \omega_{i,23} &= 0 \end{aligned}$$

A known value of the focal length  $\alpha_v$  can only be used to formulate linear equations if the other intrinsics are also known [42]. In such a fully calibrated case, other algorithms might be better suited, so we neglect that case in the following.

By substituting  $\omega_i$  in the above equations according to (16), we get the following *linear* equations on  $\mathbf{Z}$ :

$$\begin{aligned}\tau_i^2 (\mathbf{U}_i^{-\top} \mathbf{Z} \mathbf{U}_i^{-1})_{11} - (\mathbf{U}_i^{-\top} \mathbf{Z} \mathbf{U}_i^{-1})_{22} &= 0 \\ u_{i,0} (\mathbf{U}_i^{-\top} \mathbf{Z} \mathbf{U}_i^{-1})_{11} + (\mathbf{U}_i^{-\top} \mathbf{Z} \mathbf{U}_i^{-1})_{13} &= 0 \\ v_{i,0} (\mathbf{U}_i^{-\top} \mathbf{Z} \mathbf{U}_i^{-1})_{22} + (\mathbf{U}_i^{-\top} \mathbf{Z} \mathbf{U}_i^{-1})_{23} &= 0\end{aligned}$$

**Constant camera intrinsics** In the case that two cameras  $i$  and  $j$  are known to have the same, yet unknown value for one intrinsic parameter, we in general obtain *quadratic* equations on  $\mathbf{Z}$ . For example, the assumption of equal aspect ratios leads to the equation:

$$(\mathbf{U}_i^{-\top} \mathbf{Z} \mathbf{U}_i^{-1})_{11} (\mathbf{U}_j^{-\top} \mathbf{Z} \mathbf{U}_j^{-1})_{22} = (\mathbf{U}_j^{-\top} \mathbf{Z} \mathbf{U}_j^{-1})_{11} (\mathbf{U}_i^{-\top} \mathbf{Z} \mathbf{U}_i^{-1})_{22}.$$

In practice, we only use available linear equations. In some minimal cases, quadratic equations as above might be useful to find a unique solution or a finite set of solutions, if the available linear constraints are insufficient.

The situation is different if *all* intrinsic parameters of two (or more) views are known to be identical. In that case, we can obtain linear equations instead of quadratic ones, as shown in [17]: the matrices  $\mathbf{U}^i$  are scaled such as to have unit determinant. Then we can write the following component-wise matrix equality between any pair  $(i, j)$  of views:

$$\mathbf{U}_i^{-\top} \mathbf{Z} \mathbf{U}_i^{-1} - \mathbf{U}_j^{-\top} \mathbf{Z} \mathbf{U}_j^{-1} = \mathbf{0}_{3 \times 3}$$

This represents 6 *linear* equations on  $\mathbf{Z}$  for each pair of views, among which 4 are independent.

#### 4.4.2 Prior Information on Parallelepipeds

From equation (14), we have:

$$b_k \mathbf{S}_k \mathbf{L}_k = \mathbf{T} \mathbf{V}_k^{\top}$$

Due to the orthogonality of  $\mathbf{S}_k$ , we get:

$$b_k^2 \underbrace{\mathbf{L}_k^{\top} \mathbf{L}_k}_{\mu_k} = \mathbf{V}_k \underbrace{\mathbf{T}^{\top} \mathbf{T}}_{\mathbf{Z}} \mathbf{V}_k^{\top}$$

Neglecting the unknown scale factor  $b_k$ , we obtain:

$$\mu_k \sim \mathbf{V}_k \mathbf{Z} \mathbf{V}_k^{\top}.$$

Knowledge on parallelepiped intrinsics can be used in an analogous way as knowledge about camera parameters. For example suppose we know the length ratio of two parallelepiped edges  $r_{uv} = \frac{l_u}{l_v}$ . Referring to (3), we get the following *linear* equation on  $\mathbf{Z}$ :

$$r_{k,uv}^2 \mu_{k,vv} - \mu_{k,uu} = r_{k,uv}^2 (\mathbf{V}_k \mathbf{Z} \mathbf{V}_k^T)_{vv} - (\mathbf{V}_k \mathbf{Z} \mathbf{V}_k^T)_{uu} = 0.$$

Similarly, the assumption that the  $\theta_{uv}$  is a right angle, i.e.  $\cos \theta_{uv} = 0$ , leads also to a *linear equation*:

$$\mu_{k,uv} = (\mathbf{V}_k \mathbf{Z} \mathbf{V}_k^T)_{uv} = 0.$$

As for cameras, *quadratic* equations may be derived from assumptions about two or more parallelepipeds having the same, yet unknown value for some intrinsic parameter. Furthermore, two parallelepipeds having the same shape, leads to a set of *linear* equations on  $\mathbf{Z}$ . This holds even if the parallelepipeds are of different size. Knowing in addition that they are of the same size, gives an additional *linear* equation.

Currently, we only exploit constraints on individual parallelepipeds (right angles and length ratios), since they are easier to provide for the user.

#### 4.5 Complete Algorithm

1. Estimate the canonical projection matrices  $\tilde{\mathbf{X}}_{ik}$ .
2. Compute missing  $\mathbf{X}_{ik}$ .
3. Normalize the  $\mathbf{X}_{ik}$  to unit determinant.
4. Construct the measurement matrix and compute its SVD.
5. From the SVD, extract the matrices  $\mathbf{U}_i$  and  $\mathbf{V}_k$ .
6. Establish linear equation system on  $\mathbf{Z}$  based on prior knowledge of intrinsic parameters of cameras and parallelepipeds.
7. Solve the system to least squares.
8. Extract  $\mathbf{T}$  from  $\mathbf{Z}$  using Cholesky decomposition.
9. Extract the  $\mathbf{K}_i, \mathbf{R}_i, \mathbf{L}_k, \mathbf{S}_k$  from the  $\mathbf{U}_i \mathbf{T}^{-1}$  and the  $\mathbf{T} \mathbf{V}_k^T$  using e.g. QR-decomposition. Note that at this stage the  $\mathbf{L}_k$  can only be recovered up to scale, i.e. the parallelepipeds' (relative) sizes remain undetermined.

#### 4.6 Minimal cases for the linear calibration

As mentioned in the last section, all constraints provided by knowledge on the cameras and parallelepipeds can be expressed in terms of the 5 independent parameters of the matrix  $\mathbf{Z}$ . Thus, information about a total of only five intrinsic parameters of cameras or parallelepipeds

parallelepipeds		cameras	
#	constraint(s)	#	constraint(s)
0		5	<ul style="list-style-type: none"> <li>• <math>K_1: \{s, \tau, u_0, v_0\}</math> known;</li> <li>• <math>K_2: s</math> known</li> <li>• 5 cameras with known <math>s</math></li> </ul>
1	<ul style="list-style-type: none"> <li>• 1 known length ratio</li> <li>• 1 right angle</li> </ul>	4	<ul style="list-style-type: none"> <li>• 2 cameras with known <math>s</math> and <math>\tau</math></li> <li>• 4 cameras with known <math>s</math></li> </ul>
2	<ul style="list-style-type: none"> <li>• 2 right angles</li> <li>• 1 right angle and 1 known length ratio</li> </ul>	3	<ul style="list-style-type: none"> <li>• 1 camera with <math>\{s, u_0, v_0\}</math> known</li> <li>• 3 cameras with known <math>s</math></li> </ul>
3	<ul style="list-style-type: none"> <li>• 3 right angles</li> <li>• 2 right angles and 1 known length ratio</li> <li>• 1 right angle and 2 known length ratios</li> </ul>	2	<ul style="list-style-type: none"> <li>• 1 camera with known <math>s</math> and <math>\tau</math></li> <li>• 1 camera with known <math>u_0</math> and <math>v_0</math></li> <li>• 2 cameras with known <math>s</math></li> </ul>
4	<ul style="list-style-type: none"> <li>• 3 right angles and 1 known length ratio</li> <li>• 2 right angles and 2 known length ratios</li> </ul>	1	<ul style="list-style-type: none"> <li>• 1 camera with known <math>\tau</math></li> <li>• 1 camera with known <math>s</math></li> </ul>
5	<ul style="list-style-type: none"> <li>• 3 right angles and 2 length ratios</li> <li>• <math>L_0: 2</math> and <math>L_1: 3</math> right angles</li> </ul>	0	

Table 1: Some minimal cases for the linear calibration algorithm. The problem has five degree of freedom (see text). The table contains a non-exhaustive list of cases where the number of constraints on parallelepipeds and on cameras, sum up to five.

is in general sufficient to calibrate the whole system. In table 1 we give a non-exhaustive list of practical minimal cases. Note however that certain configurations, i.e. relative positioning of cameras and parallelepipeds, represent singularities, depending on the amount of prior information available. Such singularities are discussed in section 5.

## 5 Singularities

Many calibration or self-calibration algorithms are subject to more or less severe singularities, i.e. there exist situations, where the algorithm is bound to fail. Furthermore, even in situations that are not exactly singular, but close to a singularity, the results become usually very unstable. In this section, we examine the singularities for the linear calibration algorithm described above. First, we study the singularities in the case of one parallelepiped being seen by one camera. We then study some multi-view cases, where we exploit results on critical motions for classical self-calibration.

### 5.1 One parallelepiped in a single view

We have studied all possible combinations of *a priori* knowledge, on both camera and parallelepiped intrinsic parameters leading to the linear equations (see sections 4.4.1 and 4.4.2).

In the following we will sketch the methodology followed, give proofs for one sample configuration and provide the results for all configurations studied.

We first formulate the meaning of a singularity in terms of the ingredients of the calibration algorithm. The existence of a singularity in our case means exactly that equation (7) has more than one solution for  $\omega$  and  $\mu$  that conform to all available *a priori* information, i.e. that there is at least one solution that is different from the true one. It is easy to show that the existence of a singularity does not depend on the relative *position* of the camera and the parallelepiped, only on the relative *orientation* and the *a priori* knowledge on camera and parallelepiped intrinsic parameters.

Let  $K = K_k K_u$  be the true calibration matrix, and  $K' = K_k K'_u$  the estimated one (we decompose in known and unknown parts, so  $K'$  and  $K$  share of course the known part  $K_k$ ). As for parallelepipeds, a similar decomposition into known and unknown parts is not always possible. If however, we only consider constraints arising from prior knowledge about right angles, then we can decompose as above:  $L = L_u L_k$  and  $L' = L'_u L_k$  are respectively the true and estimated intrinsic parameters of a parallelepiped.

With these definitions, a singularity exists if there are solutions for (7) with  $K'_u \neq K_u$  and  $L'_u \neq L_u$ . From (7), it is easy to derive the following equality (using  $X \sim K_k K_u R L_u L_k$ ,  $\omega' \sim K'^{-T} K'^{-1}$  and  $\mu' \sim L'^T L'$ ):

$$R^T K_u^T K'^{-T} K'^{-1} K_u R \sim L_u^{-T} L'^T L'_u L_u^{-1}.$$

A singularity, as defined above, is then equivalent to the existence of matrices  $\omega'' = K_u^T K'^{-T} K'^{-1} K_u$  and  $\mu'' = L_u^{-T} L'^T L'_u L_u^{-1}$ , with  $K'_u$  and  $L'_u$  of the desired form (given by the constraints), but which are different from the identity (otherwise,  $\omega' \sim \omega$  and  $\mu' \sim \mu$ , i.e. we would look at the true solution).

Depending on the *a priori* knowledge,  $\omega''$  and  $\mu''$  have special forms (as shown in table 2 for  $\omega''$ ), independently of the actual values of the known or unknown intrinsic parameters. Hence, the configuration is singular for calibration if the relative orientation  $R$  between parallelepiped and camera is such that there are solutions  $\omega''$  and  $\mu''$  of the required special form and different from the identity, satisfying:

$$\exists(\omega'' \neq I_{3 \times 3}, \mu'' \neq I_{3 \times 3}) : R^T \omega'' R \sim \mu'' \quad (17)$$

Based on this definition, it is a rather mechanical, though sometimes tricky, task, to derive singular relative orientations. Table 3 shows all singularities for nearly all combinations of the above cases. We explain the singularities in geometrical terms, by describing the relative orientation of the parallelepiped with respect to the camera. In the following paragraphs, we give a few comments on different cases of prior knowledge on the parallelepiped.

**Three right angles, two length ratios** in this case, the Euclidean structure of the parallelepiped is completely given (up to scale), and it can be used as a classical calibration object. There are singularities proper to the use of a parallelepiped, but of course the generic singularities described in [6] apply here too.

Known camera parameters			
(A) None	(B) $\tau$	(C) $u_0, v_0$	(D) $\tau, u_0, v_0$
$\begin{pmatrix} a & 0 & d \\ 0 & b & e \\ d & e & c \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & d \\ 0 & 1 & e \\ d & e & c \end{pmatrix}$	$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & c \end{pmatrix}$

Table 2: Structure of  $\omega''$  depending on prior knowledge on intrinsic camera parameters. Structure of  $\mu''$  is similar.

Case	Conditions for singularity
A-3-1	$\mathbf{v}$ is orthogonal to the $\mathbf{x}$ or $\mathbf{y}$ camera axis
B-3-1	$\mathbf{v}$ is parallel to the optical axis
C-3-1	$\mathbf{v}$ is parallel to any of the three camera axes
D-3-1	$\mathbf{v}$ is parallel to the optical axis
A-3-0	always (3 constraints for 4 camera intrinsics)
B-3-0	any edge is parallel to the image plane
C-3-0	any edge is parallel to a camera axis
D-3-0	any edge is parallel to the optical axis
A-2-2	too difficult to describe
B-2-2	$\mathbf{v} \parallel$ image plane and $\mathbf{w} \parallel$ optical axis or image plane
C-2-2	$\mathbf{v} \parallel$ $\mathbf{x}$ or $\mathbf{y}$ axis and $\mathbf{w}$ at $45^\circ$ angle with image plane
D-2-2	$\mathbf{v} \parallel$ $\mathbf{z}$ and $\mathbf{w} \parallel$ image plane and at $45^\circ$ to both $\mathbf{x}$ and $\mathbf{y}$ never!
A-2-1	always (three constraints for four camera intrinsics)
B-2-1	$\mathbf{v}$ is parallel to the image plane
C-2-1	$\mathbf{v}$ parallel to either camera axis
	$\mathbf{v}$ and $\mathbf{w}$ are both orthogonal to the $\mathbf{x}$ camera axis
	$\mathbf{v}$ and $\mathbf{w}$ are both orthogonal to the $\mathbf{y}$ camera axis
	$\mathbf{v}$ and $\mathbf{w}$ are parallel to the image plane
D-2-1	$\mathbf{v}$ and $\mathbf{w}$ are parallel to the image plane
A-2-0	always (two constraints for four camera intrinsics)
B-2-0	always (two constraints for three camera intrinsics)
C-2-0	$\mathbf{v}$ orthogonal to the $\mathbf{x}$ or $\mathbf{y}$ camera axis or $\parallel$ image plane
D-2-0	$\mathbf{v}$ parallel to image plane or to optical axis

Table 3: Singular relative orientations for the case of one parallelepiped seen in one camera, for various combinations of prior knowledge. Cases are denoted X-Y-Z, where  $X \in \{A,B,C,D\}$  refers to table 2 and Y and Z are the number of known right angles respectively length ratios. For further explanations, see text.

**Three right angles, one length ratio (cases \*-3-1 in table 3)** in table 3,  $\mathbf{v}$  represents the direction of the parallelepiped's edges which are not "involved" in the known length ratio.



Figure 3: Examples of singular and non-singular configurations for calibration based on a parallelepiped with three right angles; (a) The parallelepiped's vertical edge is parallel to the camera's  $y$ -axis; this configuration is singular if the camera aspect ratio and principal point are not given (cf. cases B-3-0 and C-3-0 in table 3). (b) The parallelepiped's vertical edge is parallel to the image plane; this configuration is singular if the camera's principal point is not given (case B-3-0 in table 3). (c) A rotation of the camera as shown here removes the singularities of (a) and (b).

**Two right angles (cases \*-2-\* in table 3)** in this case, the parallelepiped can be visualized as built around two rectangles sharing an edge  $\mathbf{v}$  (see figure 4). The role of  $\mathbf{w}$  can be played by one of the two rectangles' edges not parallel to  $\mathbf{v}$ .

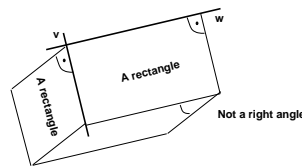


Figure 4: Illustration of the case of two right angles.

An example of the derivation of the singularities for the case C-3-0 is described in the appendix.

Several singular situations that might occur in practice, are illustrated in figure 3.

## 5.2 One parallelepiped in multiple views

Two observations are useful to characterize the singularities in the case when one parallelepiped is seen in multiple images:

- The way the canonic parallelepiped projection matrix  $X$  is computed implies that there is only an affine ambiguity in the calibration process. Thus the singularities of calibration of images viewing one parallelepiped are equivalent to singularities of generic self-calibration when the plane at infinity is known;

- It is natural to suppose that the amount of prior knowledge on the intrinsic parameters is the same for all the cameras used, thus all the matrices  $\omega_i$  belong to only one of the groups defined in table 2.

These two observations make it possible to adapt the studies on critical motions for self-calibration. In particular, we use the results presented in [22] and [23] for scenarios with known plane at infinity. Depending on the type of the knowledge about the camera (cf. table 2), the following rotations are singular for the calibration:

- A** Always critical with fewer than 5 cameras. Critical motions for 5 or more cameras are difficult to explain.
- B** Critical if the cameras' optical axes point into at most two different directions.
- C** Critical if one axis of each camera is pointing in some common direction (see figure 5).
- D** Critical if the optical axes of all cameras point into the same direction.

Results for the cases A, B, D were given in [23] and the proof of case C is given in the appendix.

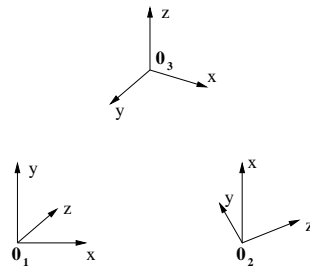


Figure 5: Illustration of critical motion for case type C (only principal point is known): one axis per camera (either optical axis  $Z$ , or pixel axes  $X$  or  $Y$ ), is pointing in some common direction.

## 6 Estimating Position and Size

In this section we propose an algorithm for the estimation of the (relative) positions of the cameras and parallelepipeds, as well as the (relative) sizes of the parallelepipeds. Consider equation (8):

$$\lambda_{ik} \tilde{\mathbf{X}}_{ik} = \mathbf{K}_i (\mathbf{R}_i \quad \mathbf{t}_i) \begin{pmatrix} \mathbf{S}_k & \mathbf{v}_k \\ \mathbf{0}^\top & 1 \end{pmatrix} \tilde{\mathbf{L}}_k$$

The leading  $3 \times 3$  sub-part of the two sides of the equation were used in section 4 to compute the intrinsic camera parameters  $\mathbf{K}_i$  and the rotation matrices  $\mathbf{R}_i$  and  $\mathbf{S}_k$ . As for the parallelepipeds' intrinsic parameters  $\mathbf{L}_k$ , they were computed up to scale only, i.e. up to the "size" of the parallelepipeds.

Let us consider this in detail. In the following, we suppose that the matrices  $\tilde{\mathbf{X}}_{ik}$  are already scaled such that the sub-matrices  $\mathbf{X}_{ik}$  have unit determinant, as in section 4, i.e.  $\lambda_{ik} = 1$ . Let  $\bar{\mathbf{K}}_i$  and  $\bar{\mathbf{L}}_k$  be the calibration matrices scaled to unit determinant. We know all matrices in the following equation:

$$\mathbf{X}_{ik} = \bar{\mathbf{K}}_i \mathbf{R}_i \mathbf{S}_k \bar{\mathbf{L}}_k$$

What we don't know is the size  $s_k$  of the parallelepipeds. Let us observe the following:

$$\tilde{\mathbf{L}}_k = \begin{pmatrix} s_k \bar{\mathbf{L}}_k & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{pmatrix} \sim \begin{pmatrix} \bar{\mathbf{L}}_k & \mathbf{0} \\ \mathbf{0}^\top & 1/s_k \end{pmatrix}$$

We may now rewrite equation (8):

$$\tilde{\mathbf{X}}_{ik} = \bar{\mathbf{K}}_i (\mathbf{R}_i \quad \mathbf{t}_i) \begin{pmatrix} \mathbf{S}_k & \mathbf{v}_k \\ \mathbf{0}^\top & 1 \end{pmatrix} \begin{pmatrix} \bar{\mathbf{L}}_k & \mathbf{0} \\ \mathbf{0}^\top & 1/s_k \end{pmatrix}$$

Let  $\mathbf{x}_{ik}$  be the fourth column of  $\tilde{\mathbf{X}}_{ik}$ . We have the following equation:

$$\mathbf{x}_{ik} = \bar{\mathbf{K}}_i (\mathbf{R}_i \quad \mathbf{t}_i) \begin{pmatrix} \mathbf{S}_k & \mathbf{v}_k \\ \mathbf{0}^\top & 1 \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ 1/s_k \end{pmatrix} = \frac{1}{s_k} \bar{\mathbf{K}}_i (\mathbf{R}_i \mathbf{v}_k + \mathbf{t}_i)$$

From this, we get an equation that is linear in all unknowns ( $s_k$ ,  $\mathbf{t}_i$  and  $\mathbf{v}_k$ ):

$$s_k \mathbf{x}_{ik} - \bar{\mathbf{K}}_i \mathbf{R}_i \mathbf{v}_k - \bar{\mathbf{K}}_i \mathbf{t}_i = \mathbf{0}$$

We may now compute the unknowns using a straightforward linear least squares problem: minimizing the sum of the squared  $L_2$  norms of the vectors on the left hand side of the above equation, over all camera-parallelepiped pairs. Note that at this stage, missing data is not an issue.

After section 4, this concludes our complete method for intrinsic and extrinsic calibration.

## 7 3D Reconstruction

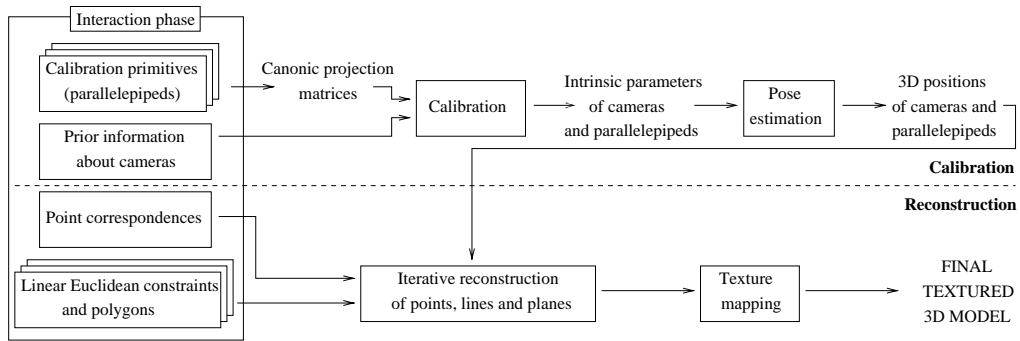


Figure 6: The calibration and reconstruction algorithms.

The presented calibration approach is well adapted to interactive 3D reconstruction from a few images. It has a major advantage over other methods: simplicity. Indeed, only a small amount of user interaction is needed for both calibration and reconstruction: a few points must be picked in the image to define the primitives' image positions. It thus seems to be an efficient and intuitive way to build models from images of any type, in particular from images taken from the Internet for which no information about the camera is known. For the reconstruction of points not belonging to the parallelepiped we use an iterative approach described in [55].

This approach is actually independent from the calibration method, although it uses the same input in the first step. Interestingly, it allows 3D models to be computed from non-overlapping photographs (see e.g. figure 18). The global scheme of our system is presented in figure 6.

The following section illustrates this approach with results obtained by solving linear systems only. Note that, in order to refine the results, non-linear optimization, taking into account prior information, might be applied.

## 8 Experimental results

### 8.1 Synthetic Scenes

Tests were performed with  $600 \times 400$  synthetic images, taken by cameras with the following intrinsic parameters:  $(\alpha_u, \alpha_v, s, u_0, v_0) = (1000, 1000, 0, 300, 200)$ . Parallelepiped parameters were varying over the different tests. The most important parameter of the experiments is the relative orientation between parallelepipeds and cameras. For a given orientation, parallelepiped vertices were projected into the images, and random Gaussian noise was added to image points. For a given setting (relative orientation, standard deviation of noise, etc.), 100 such data sets were created randomly, and used as input for calibration. Calibration was considered to have failed if any of the estimated matrices  $\omega'$  or  $\mu'$  was not positive definite (in that case,  $K'$  or  $L'$  can not be retrieved). In the following, we indicate the number of failures, as well as median values and standard deviations for estimated parameters (computed using valid calibration results only).

In the first experiment, we also compare our method with an approach based on vanishing points.

#### 8.1.1 A minimal case: 1 camera and 1 parallelepiped

This test was performed mostly to compare the calibration method described in this paper to calibration based on vanishing points in the proximity of singular configurations. To estimate vanishing points, the MLE estimator method described in [25] was used. In all tests, only 6 parallelepiped vertices are used for the calibration, which is the most common case in practice.

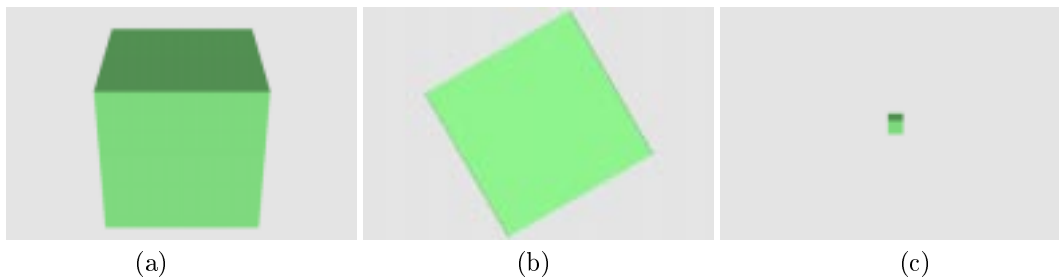


Figure 7: Parallelepiped and camera positions in the experiment 8.1.1. (a) 1st camera position; (b) maximum parallelepiped rotation; (c) minimal parallelepiped size.

**Setup** Prior information used were: the parallelepiped has only right angles and known camera parameters were  $(s, \tau) = (0, 1)$  (corresponding to case B-3-0 in section 5). This is one of the minimal cases for the calibration.

Tests were performed for different sizes and orientations of the parallelepiped with applied Gaussian noise with a standard deviation of  $\sigma=1$ . The relative parallelepiped-camera rotation varies from position  $P_0$  ( $x$  axes of parallelepiped and camera are parallel, cf. figure 7-(a)) to position  $P_n$  (the  $x$  and  $y$  axes of the parallelepiped are parallel to the image plane, cf. figure 7-(b)). According to section 5, both these positions are singular for the calibration. The maximal and minimal sizes of parallelepiped are shown on figure 7-(a) and 7-(c), respectively.

**Results** Figure 8-8 shows the number of successful calibrations, median calibration values for  $\alpha_v$  and rotation errors as functions of the rotation angle and the size of the parallelepiped. Results are shown for both the parallelepiped based approach (first row) and the vanishing point approach (second row). It can be seen that the range of positions where calibration succeeds, is about  $5^\circ$  larger on the rotation axis and 10% on the size axis for the parallelepiped-based method. Results of successful calibrations are also slightly better for the parallelepiped method.

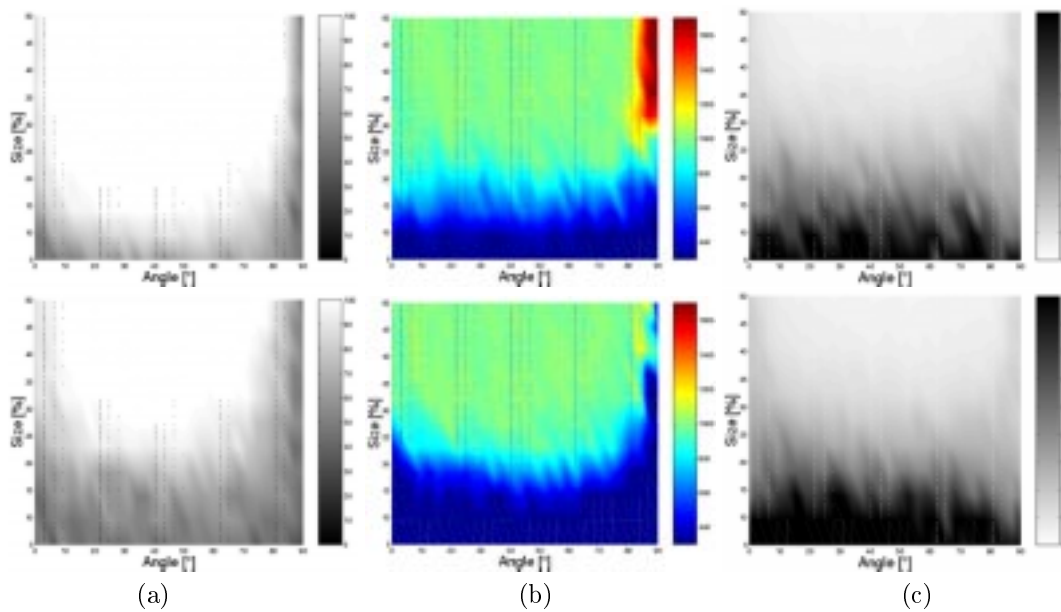


Figure 8: Calibrations results as a function of the size and relative camera-parallelepiped rotation angle. First and second row are corresponding, respectively, to parallelepiped-based and vanishing point-based calibration; Columns are corresponding to (a) Number of successful calibrations; (b) Median values of  $\alpha_v$ . Dark colors, i.e. red and blue, correspond to important calibration errors (estimated values are bigger than 150% or smaller than 50% of the true values, respectively); (c) Median rotation errors for  $\alpha_v$ .

## 8.1.2 Scenario with two cameras

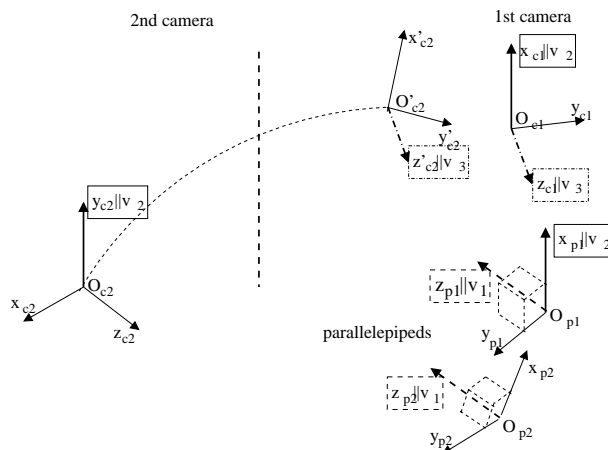


Figure 9: Experimental setup for tests on synthetic data: one camera is static ( $O_{c1}$ ) and the second one moving (from position  $O_{c2}$  to  $O'_{c2}$ ). They are watching two parallelepipeds (at  $O_{p1}$  and  $O_{p2}$ ). Parallel directions are drawn using the same line style and label.

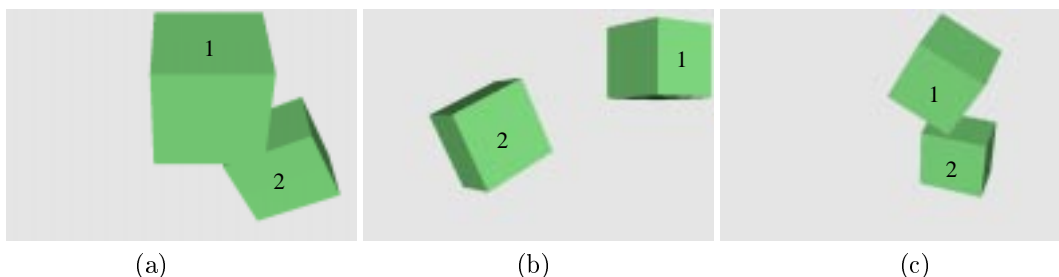


Figure 10: Three projections of parallelepipeds as seen from: (a) camera 1 ( $O_{c1}$ ); (b)-(c) camera 2 in initial and final positions ( $O_{c2}$  and  $O'_{c2}$ ) respectively.

The goal of this test was to evaluate calibration results in the proximity of singular positions, and with different numbers of calibration primitives as well as different types of prior information.

**Setup** The experimental setup is sketched in figure 9. Tests were performed using parallelepipeds with parameters  $(l_1, l_2, l_3, \theta_1, \theta_2, \theta_3) = (1, 1, 1, 90^\circ, 90^\circ, 90^\circ)$  and  $(l_1, l_2, l_3, \theta_1, \theta_2, \theta_3) = (4, 2, 3, 90^\circ, 90^\circ, 90^\circ)$ . Two cameras with zero skew and known principal points (case C, cf. section 5.2) are watching these parallelepipeds. The first camera

is static and its  $x$ -axis is parallel to the first parallelepiped's  $x$ -axis. The second camera is rotating from the position  $P_0$ , where its  $y$ -axis is parallel to the first parallelepiped's  $x$ -axis (and thus parallel to the first camera's  $x$ -axis) to the position  $P_n$  where its  $z$ -axis is parallel to the first camera's  $z$ -axis. As explained below, both positions  $P_0$  and  $P_n$  correspond to singular configurations.

All charts illustrating the results are separated into two parts, depending on the smallest relative angle between camera axes, or equivalently, the proximity to a singular position. The first part shows errors as a function of the angle between the first camera's  $y$ -axis and the 2nd camera's  $x$ -axis (these two axes being parallel in the initial position  $P_0$ ). The second part shows results as a function of the angle between the  $z$ -axes of both cameras (parallel in the final position  $P_n$ ). Labels (F-1) and (F-2) correspond to calibration using the factorization method based on 1 and 2 parallelepipeds. Here, white Gaussian noise with a  $\sigma = 1$  pixel standard deviation was added to image point positions.

**Singular configurations** Using no prior information about the parallelepipeds is equivalent to doing self-calibration of our 2 cameras, with a known plane at infinity<sup>4</sup> According to section 5.2, singularities for the camera type C occur when two of the cameras' axes (optical axes or pixel axes) are parallel. Thus, both positions  $P_0$  and  $P_n$  are singular. Note however that focal lengths  $\alpha_{v,1}$  of the first camera, and  $\alpha_{u,2}$  of the second camera can be determined in  $P_0$ , and that camera aspect ratios  $\tau_1$  and  $\tau_2$  of both cameras can be determined in  $P_n$  (see appendix). The charts 11–13, showing the intrinsic and extrinsic calibration results, reflects clearly the proximities of the singular positions.

**Intrinsic calibration** Figure 11–(a) shows the number of successful calibrations. It can be seen that at both initial and final positions  $P_0$  and  $P_n$ , calibration is very unstable, as expected.

Singular positions are stressed also in charts on figure 12. However, when the minimal angle between camera axes is larger than  $15^\circ$  the method can be considered as stable. All calibrations are successful, the relative error on obtained median values is not bigger than 7% and standard deviations are not larger than 15%. Naturally, calibration is stabler with 2 parallelepipeds than with a single one.

**Extrinsic calibration** Figure 13 presents the results for the estimation of the camera's rotation and translation. Extrinsic calibration was performed only when the previous internal calibration was successful. Displayed error values on recovered rotations (figure 13–(a)) are the angle of the relative rotation between the true  $R$  and the estimated  $R'$  (the presented values are computed as the mean rotation error for both cameras). Translation error (figure 13–(b)) is represented by the mean of the two angles between the true and estimated relative camera–parallelepiped position vectors. This allows to evaluate the geometry of the reconstructed cameras independently from errors caused by global scaling and errors on the

<sup>4</sup>Note that the presence of a second parallelepiped, as long as its Euclidean shape is completely unknown, does not remove any calibration ambiguity.

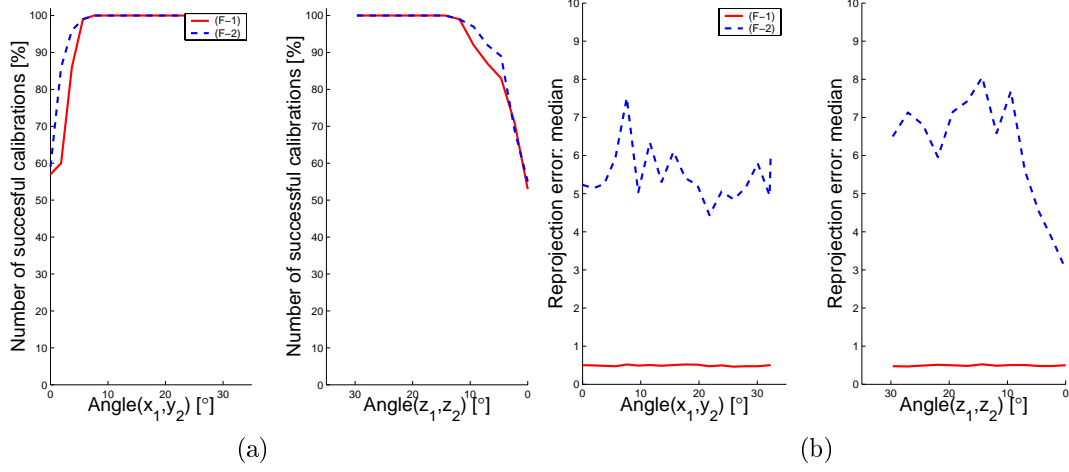


Figure 11: (a) Number of successful calibrations; (b) Median values of reprojection errors. Plots described by (F-1) and (F-2) correspond to calibration based on 1 and 2 parallel epipeds.

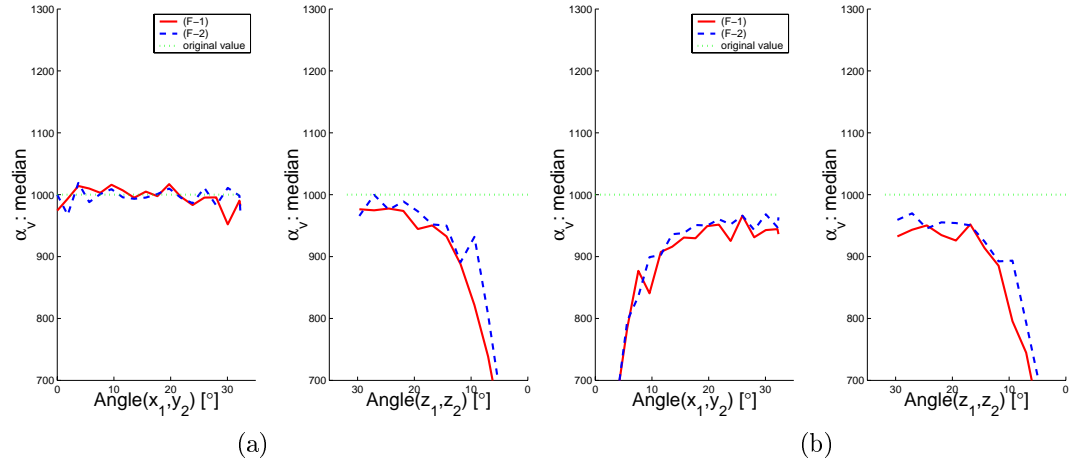


Figure 12: Median values obtained for the focal length of (a) first camera; (b) second camera. Notations are similar to figure 11.

rotation. For a standard deviation of  $\sigma = 1$  and a minimal angle of  $10^\circ$  between the camera axes, the rotation error is never larger than  $10^\circ$ , with a  $5^\circ$  median value, and the translation error is similar.

As expected, the rotation and translation estimation is most stable and accurate with the factorization method using two parallelepipeds. However, increasing the number of scene constraints (i.e. number of calibration primitives) increases also the reprojection error. The increase of the reprojection error for the methods based on two calibration parallelepipeds can be seen in chart 11–(b).

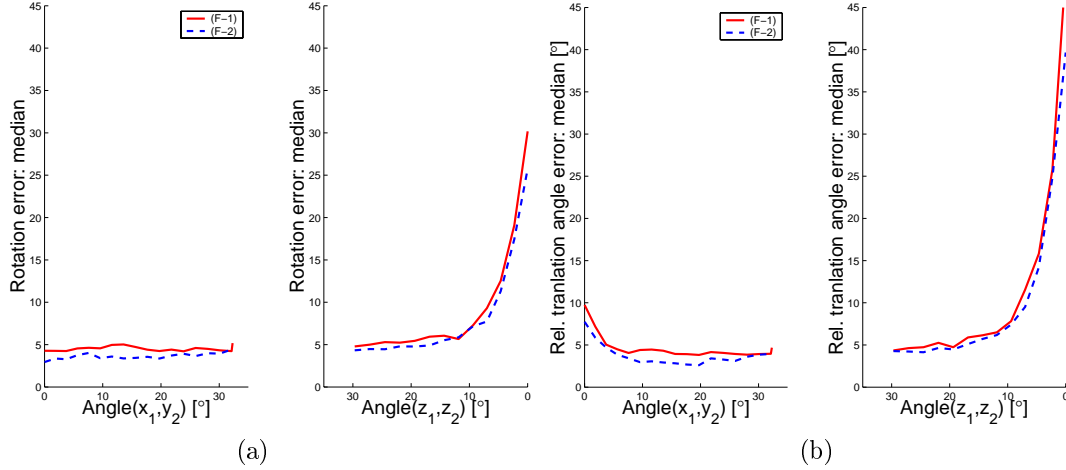


Figure 13: Extrinsic calibration results. (a) Median values of the rotation error; (b) Median values of the angle between the true and the estimated direction between the camera–parallelepiped relative translation vectors. Notations are similar to figure 11.

**Using knowledge on parallelepiped angles** Additional information about the parallelepiped’s structure decreases the number of singularities. The same experiment as above is considered here, but now we use information about right angles in the parallelepipeds (cf. case C-3-0 in section 5). Our theoretical results are confirmed: singularities only occur when each of the parallelepipeds and cameras has one axis parallel to one common direction. This is the case of position  $P_0$  for the calibration based on the first parallelepiped, where the estimation of  $\alpha_{v,2}$  remains ambiguous. Results are shown on figure 14.

## 8.2 Results on real scenes

In this section, we present 3D reconstruction results of our method when applied to indoor and outdoor scenes. These examples correspond to situations where automatic methods are bound to fail: small sets of images are used and occlusions frequently occur. Each reconstruction was performed in two steps: first, one or more parallelepipeds were used to calibrate the intrinsic and extrinsic camera parameters; second, scene points and geometrical

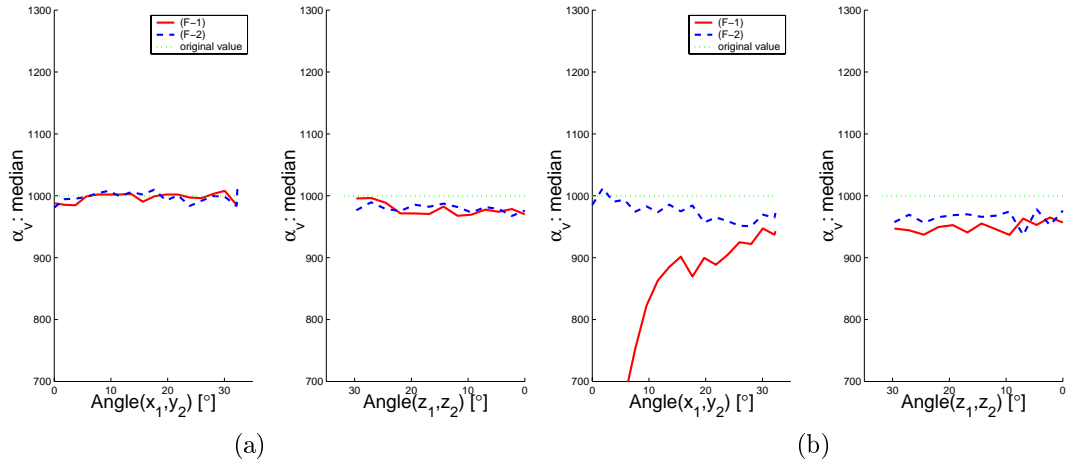


Figure 14: Calibration results for the focal length of the (a) first; (b) second camera. 1 and 2 parallelepipeds with known right angles were used.

constraints were used for the reconstruction (see sec. (7)). Results from both single and multiple images are shown.

### 8.2.1 Bedroom scene

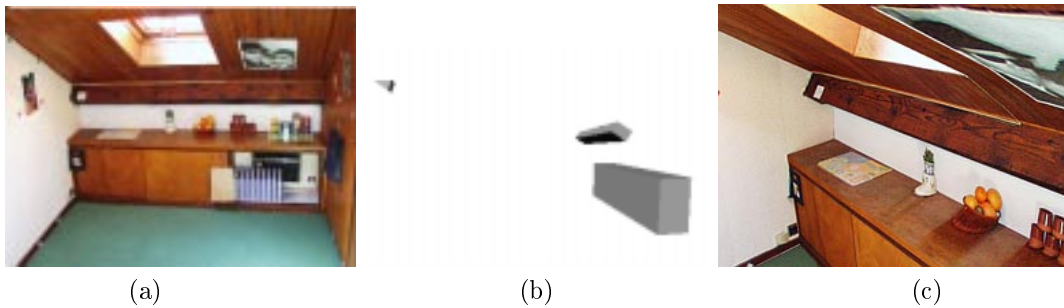


Figure 15: Bedroom scene: (a) The original image; (b) The parallelepipeds present in the model; (c) The textured model seen from a different viewpoint.

Figure (15–(a)) shows the original image used for the reconstruction of a bedroom scene. Such image was taken with a very small camera focal length, leading therefore to a slight optical distortion which was not corrected here. The calibration was based on the cupboard in the central part of the image. Note that the camera was almost frontoparallel with respect

to the cupboard, and thus close to a singular situation for calibration. The prior informations used for the calibration were: right parallelepiped angles, null camera skew parameter and central image principal point. The full model was reconstructed using 29 points, constrained by 2 parallelepipeds (the cupboard and the wooden belt), 3 parallelograms and 6 colinearity and coplanarity constraints. Figure (15-(b)) shows the reconstructed parallelepipeds as well as the camera pose, and figure (15-(c)) shows the scene rendered from a new viewpoint.

### 8.2.2 Notre-Dame square scene

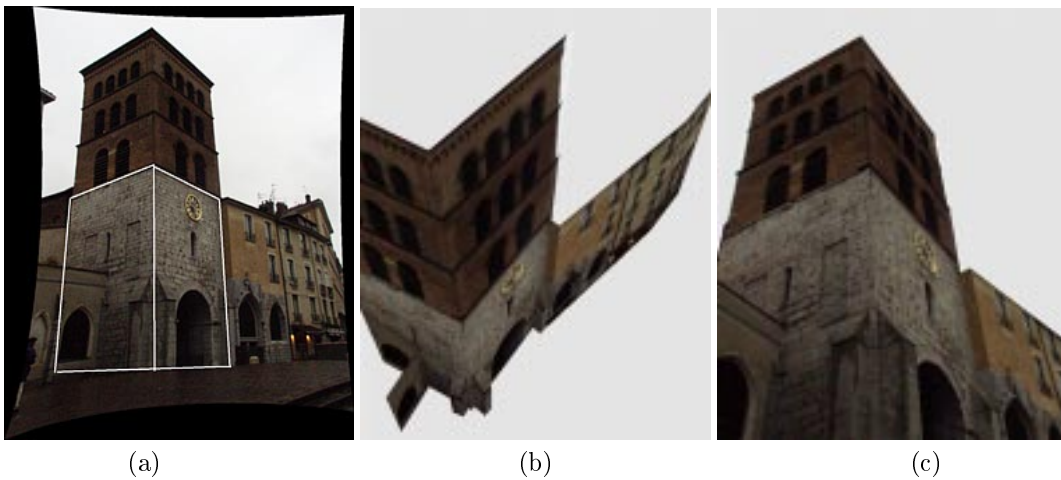


Figure 16: Notre-Dame square scene: (a) the original image; (b)-(c) screen-shots of the model obtained using image (a) only.

In this section, we present reconstructions of Notre-Dame square in Grenoble from single and multiple images. In both cases, the image distortion was corrected in a preliminary step.

**Reconstruction from one image** The image and the calibration primitive for the single image reconstruction are shown in figure (16-(a)). The prior information used for the calibration were: right parallelepiped angles, null camera skew parameter and central image principal point. The final model is composed of 42 points, 3 parallelepipeds, 4 parallelograms and 4 lines and planes. New viewpoints of the model are shown in images (16-(b)(c)).

**Reconstruction from multiple images** The sequence used for the reconstruction is composed of 15 images whose sizes vary from  $768 \times 1024$  to  $960 \times 1280$  pixels. The calibration was based on 3 parallelepipeds (shown in figures (17-(a)(d))). The prior informations

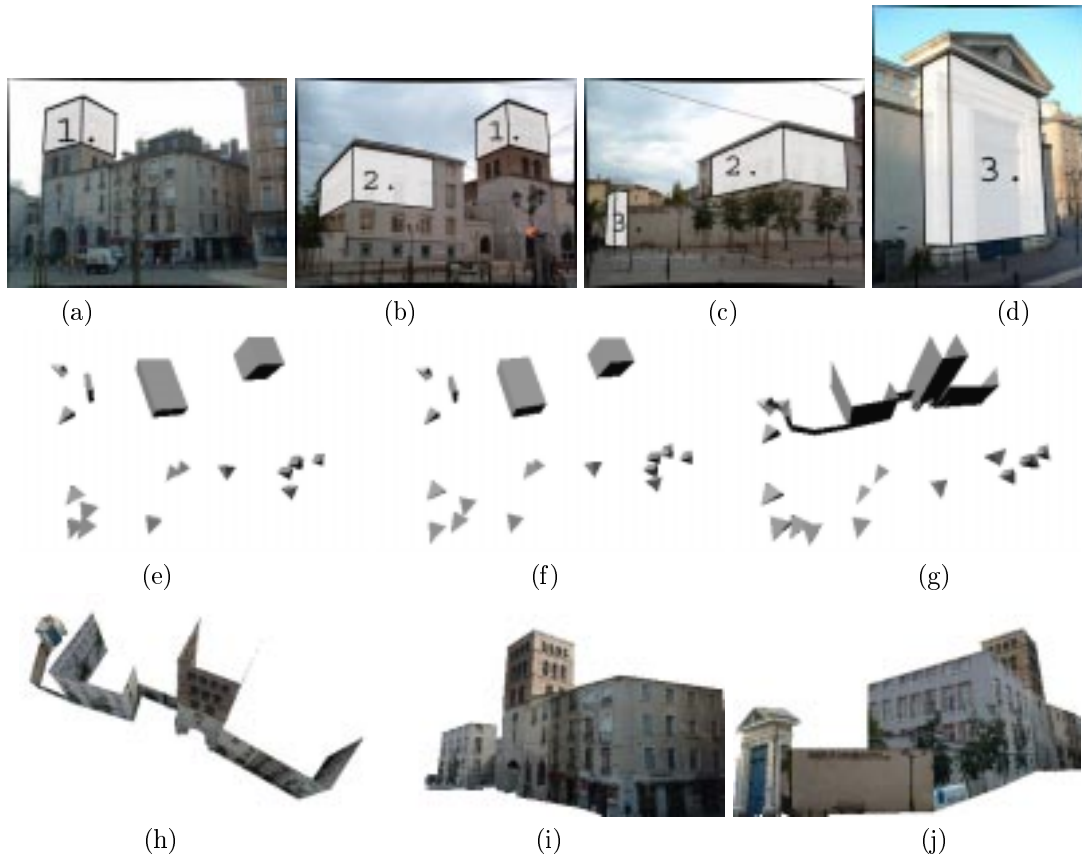


Figure 17: Notre-Dame square scene: (a)–(d) 4 images from the 15 used for the reconstruction. Parallelepipeds used for the reconstruction are white painted; (e) Cameras and parallelepipeds as estimated by the proposed linear factorization method; (f) Cameras and parallelepipeds parameters optimized by a non-linear method; (g) Cameras and 194 model points optimized by an unconstrained non-linear method; (h)–(j) New viewpoints of the textured model.

used for the calibration were: right angles for parallelepipeds 1 and 2, null skew parameters, unit aspect ratios and central principal points for all images. Parallelepiped 3 is relatively small in the images where both parallelepipeds 2 and 3 appear. Consequently, the estimation of its vertices is unstable, and thus its parameters were not used for the calibration. The calibration was performed in two steps: first, the proposed linear factorization approach is applied, and second, the parameters obtained from the previous step are optimized. Before the second step, the mean reprojection error of parallelepipeds' vertices used for the

calibration was 28 pixels. The errors occurred mainly in the images that were calibrated using parallelepiped 3. The non-linear optimization step reduced dramatically this error to 3 pixels.

After the calibration step, primitives were added and reconstructed so that the final model is composed of 194 points, 19 planes and 25 lines. The mean reprojection error over all the model points increased up to 8 pixels. As might be expected, the most important errors always occurred in images calibrated using parallelepiped 3.

For comparison, a global bundle adjustment was performed over all the model points and the camera focal lengths. This reduced the reprojection error to 2 pixels. It did not reduce, however, the small artifacts occurring in the final model. Both non-linear methods were implemented without code optimization and derivatives were computed numerically. The parallelepiped-based optimization converged after 24 iterations and 15 seconds, and the standard point-based optimization converged after 51 iterations and 20 minutes (on a Pentium III, 733 Mhz).

The calibration primitives and the cameras reconstructed using the factorization method, the parallelepiped-based non-linear optimization and the point-based non-linear optimization are shown, respectively, in figures 17-(e-g). Synthetic views of the scene reconstructed using the parallelepiped-based calibration are shown in images 17-(h-j).

### 8.2.3 Opposite viewpoint scene

Figure 18 shows the reconstruction of a modern building from 2 images taken from opposite viewpoints. The parallelepiped used for calibration and the estimated cameras' positions are shown in the two original images (see figures 18-(a)(b)). In the first image, intersections of lines were computed to obtain the six points required to define a parallelepiped (see figure 18-(a)). The parallelepiped and the cameras reconstructed by the factorization algorithm are shown in the image 18-(c). New viewpoints of the whole model, composed of 32 points, 13 parallelograms and 6 planes are shown in the figures (18-(d)(f)).

### 8.2.4 Kio towers

The Kio towers reconstruction was based on 3 images and 2 calibration primitives. One of the images used for the reconstruction is shown in figure (19-(a)). The information used for the calibration were: 2 right angles in each tower, null camera skew parameters, unit aspect ratios and central principal points. The cameras and primitives reconstructed are shown in figure (19-(b)) and a new viewpoint of the scene is shown in figure (19-(c)).

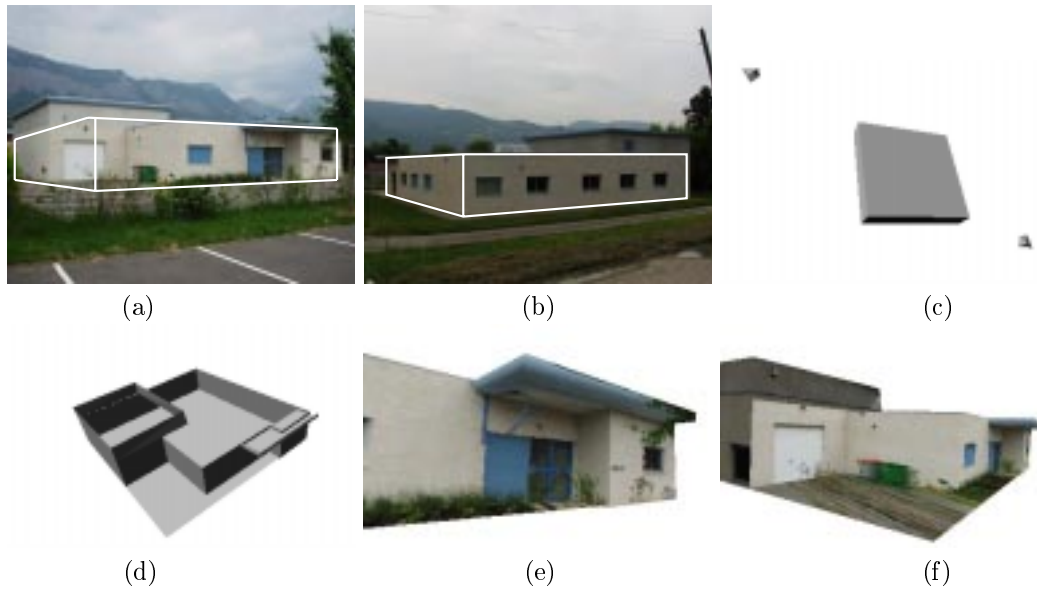


Figure 18: Opposite viewpoints scene: (a),(b)the original images used for the reconstruction; (c) the reconstruction scenario with the computed model and the cameras' positions; (d),(e) New viewpoints of the model.

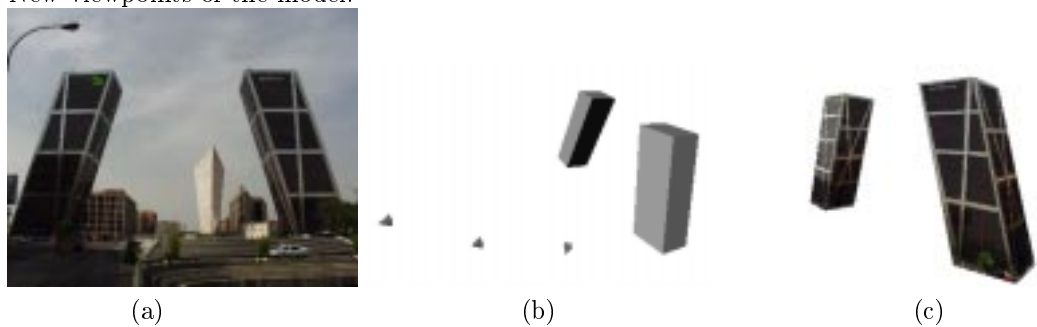


Figure 19: Kio towers in Madrid: (a) the original image, (b)(c) screen-shots of the reconstructed model.

## 9 Conclusion

We have presented an approach for calibration, pose estimation and 3D model acquisition from several uncalibrated images based on user-provided geometric constraints on the scene. Useful constraints such as parallelism, coplanarity and right angles, can often be nicely modeled via parallelepipeds. Especially, this allows to couple together constraints between

several neighboring scene primitives (points, lines, planes), which potentially brings about a higher stability than only using constraints between pairs of primitives. The projections of parallelepipeds already encode the affine structure of the scene. Metric information (length ratios and angles) is then combined with prior information on camera parameters in a self-calibration type approach, performing complete calibration and pose estimation. This is formulated in a factorization framework. The usual problems of missing data and unknown scale factors are dealt with relatively easily, and a satisfying solution can be obtained with already a small number of images and correspondences (starting from 4 correspondences per image pair or 6 per image and parallelepiped).

A detailed study on singular cases of this approach is also presented: singularities are derived theoretically, and the impact on the method's performance due to the proximity to singular configurations is shown by simulated experiments. Experiments with real images show that our calibration approach gives excellent initial results for general 3D model reconstruction methods.

We believe that an approach such as the one presented here, is a useful tool for easily calibrating cameras using images of unknown though constrained scenes. Also, it allows to efficiently obtain models of the global structure of scenes (including camera pose), which are good starting points for more automatic reconstruction methods.

## A Proof of the singularities for the calibration of cameras with known skew and principal point

In this section we derive the relative rotation matrices that are singular for case C (cf. table 2): parallelepipeds are observed by a camera with known skew and principal point. Two sub-cases will be considered: one camera viewing a parallelepiped with only right angles and multiple cameras viewing one parallelepiped with unknown shape.

Let  $\mathbf{e}_1 = (1, 0, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0)$  and  $\mathbf{e}_3 = (0, 0, 1)$  be the usual three *elementary vectors*. We now prove the following proposition, that will be used below to derive singularities.

**Proposition 1** *Let matrices  $\mathbf{A} = \text{diag}(a_1, a_2, a_3)$  and  $\mathbf{B} = \text{diag}(b_1, b_2, b_3)$  be diagonal and matrix  $\mathbf{R}$  be a rotation matrix. There exist matrices  $\mathbf{A}$  and  $\mathbf{B}$  different from  $\mathbb{1}_{3 \times 3}$  satisfying*

$$\mathbf{A} \sim \mathbf{R}^\top \mathbf{B} \mathbf{R} \quad (18)$$

*in two following cases:*

- (i) *all columns  $\mathbf{r}_j$  of  $\mathbf{R}$  are elementary vectors (up to sign), i.e. there exist permutations  $(j_1, j_2, j_3)$  and  $(i_1, i_2, i_3)$  of  $(1, 2, 3)$  with:*

$$\forall k \in \{1, 2, 3\} : \mathbf{r}_{j_k} = \pm \mathbf{e}_{i_k}$$

- (ii) *a single column  $j_1$  of  $\mathbf{R}$  is an elementary vector (up to sign):  $\mathbf{r}_{j_1} = \pm \mathbf{e}_{i_1}$ . In such a case matrices  $\mathbf{A}$  and  $\mathbf{B}$  must satisfy  $a_{j_2} = a_{j_3}$  and  $b_{i_2} = b_{i_3}$ , where  $\{j_2, j_3\} = \{1, 2, 3\} \setminus \{j_1\}$  and  $\{i_2, i_3\} = \{1, 2, 3\} \setminus \{i_1\}$ .*

Note that in case (ii) there is no ambiguity in the estimation of values  $\frac{a_{j_2}}{a_{j_3}}$  and  $\frac{b_{i_2}}{b_{i_3}}$ .

*Proof:* Consider matrices  $\mathbf{A}$  and  $\mathbf{B}$  that satisfy equation (18), and let  $k$  be the scale factor such that:  $\mathbf{A} = k\mathbf{R}^\top \mathbf{B} \mathbf{R}$ . It follows that  $\mathbf{A}$  and  $\mathbf{B}$  are similar and that their spectra are identical up to the scale factor  $k$ , i.e.  $(a_1, a_2, a_3) = kP(b_1, b_2, b_3)$ , where  $P$  may denote any permutation [15].

Equation (18) can be rewritten as:

$$\mathbf{R} \begin{pmatrix} a_1 & & \\ & a_2 & \\ & & a_3 \end{pmatrix} = k \begin{pmatrix} b_1 & & \\ & b_2 & \\ & & b_3 \end{pmatrix} \mathbf{R} \quad (19)$$

In the following the elements of matrix  $\mathbf{R}$  will be denoted as  $r_{ij}$ .

" $\Rightarrow$ " Let us begin by multiplying equation (19), successively, by  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ . The following equations are obtained:

$$\forall i, j \in \{1..3\} : (a_j - kb_i)r_{ij} = 0. \quad (20)$$

The condition  $(a_1, a_2, a_3) = kP(b_1, b_2, b_3)$  can be satisfied in six possible ways, corresponding to the six permutations of elements  $(b_1, b_2, b_3)$ . In the following let us consider, without loss of generality, that  $(a_1, a_2, a_3) = k(b_1, b_2, b_3)$ . The reasoning for the remaining permutations can be done analogously.

Let us consider separately the three following cases:

- (i) the  $a_i$  are mutually different.

According to the assumptions:

$$\left\{ \begin{array}{l} (a_1 \neq kb_2) \\ (a_1 \neq kb_3) \\ (a_2 \neq kb_1) \\ (a_2 \neq kb_3) \\ (a_3 \neq kb_1) \\ (a_3 \neq kb_2) \end{array} \right. \Rightarrow \left\{ \begin{array}{l} (r_{21} = 0) \\ (r_{31} = 0) \\ (r_{12} = 0) \\ (r_{32} = 0) \\ (r_{13} = 0) \\ (r_{23} = 0) \end{array} \right.$$

$$\begin{aligned} \mathbf{r}_j^\top \mathbf{r}_j &= 1 \\ \det \mathbf{R} &= 1. \end{aligned} \tag{21}$$

All the conditions (21) are satisfied if and only if the matrix  $\mathbf{R}$  is of the form

$$\mathbf{R} = \begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix}$$

such that  $\det \mathbf{R} = 1$ . Thus,  $\mathbf{R}$  is of form (i) given in proposition 1.

Performing the analogous reasoning for the other permutations, it is easy to show that all singular cases for the assumptions made here, are described by case (i) of the proposition.

- (ii)  $a_1 = a_2 \neq a_3$ .

According to the assumptions:

$$\left\{ \begin{array}{l} (a_1 \neq kb_3) \\ (a_2 \neq kb_3) \\ (a_3 \neq kb_1) \\ (a_3 \neq kb_2) \end{array} \right. \Rightarrow \left\{ \begin{array}{l} (r_{31} = 0) \\ (r_{32} = 0) \\ (r_{13} = 0) \\ (r_{23} = 0) \end{array} \right.$$

$$\begin{aligned} \mathbf{r}_j^\top \mathbf{r}_j &= 1 \\ \det \mathbf{R} &= 1. \end{aligned} \tag{22}$$

The conditions (22) imply that  $\mathbf{r}_3 = \pm \mathbf{e}_3$ . Together with the assumption  $a_1 = a_2$  (which implies  $b_1 = b_2$ ), we identify here the conditions corresponding to case (ii) of the proposition. Again, applying the same reasoning to all possible cases of  $a_{j_1} = a_{j_2}$  and all permutation  $P$  in  $(a_1, a_2, a_3) = kP(b_1, b_2, b_3)$ , leads always to case (ii) of the proposition. Note that all matrices  $\mathbf{R}$  satisfying condition (i) satisfy also condition (ii).

(iii)  $a_1 = a_2 = a_3$ .

This is equivalent to  $A \sim B \sim I_{3 \times 3}$ , which is excluded in the proposition.

" $\Leftarrow$ " Let  $(i_1, i_2, i_3)$  and  $(j_1, j_2, j_3)$  be two permutations of  $(1, 2, 3)$  and the rotation matrix  $R$  such that one of its columns is an elementary vector:  $\mathbf{r}_{j_1} = \mathbf{e}_{i_1}$ . Applying the orthogonality condition on  $R$ , its elements  $r_{ij}$  must satisfy, for some angle  $\theta$ :

$$\begin{cases} a_{j_1} &= \pm kb_{i_1} \\ a_{j_2} \cos \theta &= kb_{i_2} \cos \theta \\ a_{j_2} \sin \theta &= kb_{i_3} \sin \theta \\ a_{j_3} \sin \theta &= kb_{i_2} \sin \theta \\ a_{j_3} \cos \theta &= kb_{i_3} \cos \theta \end{cases} \quad (23)$$

Equation (19) implies that all elements  $r_{ij}$  must satisfy  $r_{ij}a_j = kr_{ij}b_i$ . Developing this expression for  $i = i_1, i_2, i_3$  and  $j = j_1, j_2, j_3$  and applying conditions (24) leads to:

$$\begin{cases} r_{i_1 j_1} = \pm 1 \\ r_{i_1 j_2} = r_{i_1 j_3} = r_{i_2 j_1} = r_{i_3 j_1} = 0 \\ r_{i_2 j_2} = r_{i_3 j_3} = \cos \theta \\ r_{i_3 j_2} = -r_{i_2 j_3} = -\sin \theta \end{cases} \quad (24)$$

Conditions (23) are satisfied if and only if:

$$\begin{aligned} &(\sin \theta = 0 \quad \wedge \quad a_j, b_i \in \mathbb{R}) \\ &\vee (\cos \theta = 0 \quad \wedge \quad a_j, b_i \in \mathbb{R}) \\ &\vee (\theta \in \{0..2\pi\} \quad \wedge \quad a_{j_2} = a_{j_3} = kb_{i_2} = kb_{i_3}) \end{aligned}$$

Thus, equation (18) is satisfied for the following forms of matrices  $A, B, R$ :

1. All columns  $\mathbf{r}_j = \mathbf{e}_i$  and the values  $a_j, b_i$  are arbitrary. This corresponds to case (i) of the proposition.
2. One column  $\mathbf{r}_{j_1} = \mathbf{e}_{i_1}$  and the values  $a_{j_2} = a_{j_3} = kb_{i_2} = kb_{i_3}$ . Only matrices with two equal diagonal elements satisfy equation (18). This means that there is no ambiguity in the computation of ratios of elements  $\frac{a_{j_2}}{a_{j_3}}$  and  $\frac{b_{i_2}}{b_{i_3}}$ . This corresponds to case (ii) of the proposition. ■

## A.1 1 parallelepiped seen by 1 camera

In the following, the rotation matrices that are singular for case C-3-0 (cf. section 5) will be derived. Case C-3-0 corresponds to a known principal point and skew and parallelepiped

with three right angles, i.e.  $\omega''$  and  $\mu''$  are diagonal matrices:  $\omega'' = \text{diag}(b_1, b_2, b_3)$  and  $\mu'' = \text{diag}(a_1, a_2, a_3)$ .

According to the definition of singularity given in section 5, proposition 1 gives us directly the singular relative rotations R:

- (i)  $R = (\pm \mathbf{e}_{i_1}, \pm \mathbf{e}_{i_2}, \pm \mathbf{e}_{i_3})$ , where  $(i_1, i_2, i_3)$  is a permutation of  $(1, 2, 3)$  and such that  $\det R = +1$ . In that case, none of the elements of  $\omega''$  and  $\mu''$  can be estimated, i.e. calibration fails completely. This case corresponds to configurations where each edge of the parallelepiped is parallel to one of the three camera axes.
- (ii) One of the columns of R is an elementary vector. This corresponds to the case when one camera axis is parallel to one axis of the parallelepiped. In that case, although calibration fails overall, some individual parameters may indeed be estimated successfully. According to the results of proposition (1) and to the form of matrices  $\omega$  and  $\mu$ , we can summarize this in the following table. For example, if  $\mathbf{r}_2 = \pm \mathbf{e}_3$ , the aspect ratio  $\tau$  as well as the length ratio  $\frac{l_1}{l_3}$  of parallelepiped edges, can be estimated.

$= \pm$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{r}_1$	$\frac{l_2}{l_3}, \alpha_v$	$\frac{l_2}{l_3}, \alpha_u$	$\frac{l_2}{l_3}, \tau$
$\mathbf{r}_2$	$\frac{l_1}{l_3}, \alpha_v$	$\frac{l_1}{l_3}, \alpha_u$	$\frac{l_1}{l_3}, \tau$
$\mathbf{r}_3$	$\frac{l_1}{l_2}, \alpha_v$	$\frac{l_1}{l_2}, \alpha_u$	$\frac{l_1}{l_2}, \tau$

## A.2 1 parallelepiped seen in 2 cameras

As mentioned previously, due to the observation of a parallelepiped, we know affine scene structure, i.e. the plane at infinity in a projective reconstruction. We can thus apply the appropriate results of [22] on singularities for general self-calibration. We will show that self-calibration of cameras with known principal point and zero skew is singular for the same type of rotations as the calibration in case C-3-0 described above. Let the true intrinsic parameters of the cameras be represented by diagonal matrices  $\omega_1$  and  $\omega_2$  and the estimated matrices be represented by diagonal matrices  $\omega'_1, \omega'_2$ . The coordinate system is defined such that the rotation of the first camera is  $R_1 = I_{3 \times 3}$  and the rotation of the second camera is  $R_2$ . Let  $C_f$  denote a  $3 \times 3$  matrix corresponding to a false absolute conic on the plane at infinity. The image of a conic lying on the plane at infinity is [22] represented by

$$\omega'_i \sim K_i^{-T} R_i C_f R_i^T K_i^T, \quad (25)$$

Let us consider matrices  $\omega''_i = K_i^T \omega'_i K_i^{-T}$ . Then:

$$\omega''_i \sim R_i C_f R_i^T. \quad (26)$$

Matrices  $\omega''_i$  must be diagonal. Let us represent them as  $\omega''_1 \sim \text{diag}(a_1, a_2, a_3)$  and  $\omega''_2 \sim \text{diag}(b_1, b_2, b_3)$ . Using the equation (26) and due to  $R_1 = I_{3 \times 3}$

we have  $C_f \sim \text{diag}(a_1, a_2, a_3)$ . Equation (26) applied to the second camera is  $\omega_2'' \sim R_2 C_f R_2^\top$ . The rotation is singular when there exist matrices  $C_f$  and  $\omega_2''$  different from the identity that satisfy equation (26). According to proposition 1 the following cases are singular:

- (i)  $R_2 = (\pm \mathbf{e}_{i_1}, \pm \mathbf{e}_{i_2}, \pm \mathbf{e}_{i_3})$ , where  $(i_1, i_2, i_3)$  is a permutation of  $(1, 2, 3)$  and  $\det R_2 = +1$ . No intrinsic parameter can be estimated. This corresponds to cases where each axis of the first camera is parallel to some axis of the second camera (not necessarily to the corresponding axis).
- (ii) One column of  $R_2$  is an elementary vector. This corresponds to cases where one axis of the first camera is parallel to one axis of the second camera. Using the results of proposition 1 and the form of matrices  $\omega_1$  and  $\omega_2$ , in the following special cases some individual intrinsic parameters can be estimated:

$= \pm$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{r}_1$	$\alpha_{v,1}, \alpha_{v,2}$	$\alpha_{v,1}, \alpha_{u,2}$	$\alpha_{v,1}, \tau_2$
$\mathbf{r}_2$	$\alpha_{u,1}, \alpha_{v,2}$	$\alpha_{u,1}, \alpha_{u,2}$	$\alpha_{u,1}, \tau_2$
$\mathbf{r}_3$	$\tau_1, \alpha_{v,2}$	$\tau_1, \alpha_{u,2}$	$\tau_1, \tau_2$

## References

- [1] M. Armstrong, A. Zisserman, and P. Beardsley. Euclidean structure from uncalibrated images. In E. Hancock, editor, *Proceedings of the fifth British Machine Vision Conference, York, England*, volume 2, pages 509–518, September 1994.
- [2] P. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 683–695. Springer-Verlag, April 1996.
- [3] D. Bondyfalat and S. Bougnoux. Imposing euclidean constraints during self-calibration processes. In R. Koch and L. Van Gool, editors, *Proceedings of the SMILE Workshop on 3D Structure from Multiple Images of Large-Scale Environments, Freiburg, Germany*, volume 1506 of *Lecture Notes in Computer Science*, pages 224–235. Springer-Verlag, June 1998.
- [4] D. Bondyfalat, T. Papadopoulo, and B. Mourrain. Using scene constraints during the calibration procedure. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, pages 124–130, July 2001.
- [5] T. Buchanan. The twisted cubic and camera calibration. *Computer Vision, Graphics and Image Processing*, 42(1):130–132, April 1988.

- 
- [6] B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4:127–140, 1990.
  - [7] C.S. Chen, C.G. Yu, and Y.P. Hung. New calibration-free approach for augmented reality based on parameterized cuboid structure. *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, 4:30–37, 1999.
  - [8] R. Cipolla and E. Boyer. 3d model acquisition from uncalibrated images. In *Proceedings of IAPR Workshop on Computer Vision, Chiba, Japan*, pages 559–568, November 1998.
  - [9] A. Criminisi, I. D. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000.
  - [10] L. de Agapito, R.I. Hartley, and E. Hayman. Linear self-calibration of a rotating and zooming camera. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, 1999.
  - [11] P.E. Debevec, C.J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: a hybrid geometry-and image-based approach. In *SIGGRAPH '96, New Orleans*, August 1996.
  - [12] A.R. Dick, P.H.S. Torr, S.F. Ruffle, and R. Cipolla. Combining single view recognition and multiple view stereo for architectural scenes. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, 2001.
  - [13] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.
  - [14] A.W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *European Conference on Computer Vision*, pages 311–326, June 1998.
  - [15] G.H. Golub and C.F. van Loan. *Matrix Computation*. The Johns Hopkins University Press, Baltimore, 1989.
  - [16] E. Grossmann and J. S. Victor. Single and multi-view reconstruction of structured scenes. In *Proceedings of the Fifth Asian Conference on Computer Vision, Melbourne, Australia*, pages 93–104, January 2002.
  - [17] R. I. Hartley. Self-calibration of stationary cameras. *International Journal of Computer Vision*, 22(1):5–23, 1997.
  - [18] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proceeding of the DARPA-ESPRIT workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 187–202, October 1993.

- 
- [19] R.I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.
- [20] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, June 2000.
- [21] S. Heuel. Points, lines and planes and their optimal estimation. In *Pattern Recognition, 23rd DAGM Symposium*, number 2191 in LNCS, pages 92–99. Springer, September 2001.
- [22] F. Kahl. Critical motions and ambiguous euclidean reconstructions in auto-calibration. In *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, volume 2, pages 469–475, 1999.
- [23] F. Kahl, B. Triggs, and K. Åström. Critical motions for auto-calibration when some intrinsic parameters can vary. *Journal of Mathematical Imaging and Vision*, 13(2):131–146, October 2000.
- [24] J. Kosecka and W. Zhang. Video compass. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, pages 476–491, May 2002.
- [25] D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, USA*, pages 482–488, June 1998.
- [26] D. Liebowitz and A. Zisserman. Combining scene and auto-calibration constraints. In *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, September 1999.
- [27] E. Malis and R. Cipolla. Camera self-calibration from unknown planar structures enforcing the multi-view constraints between collineations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(9), September 2002.
- [28] D. Martinec and T. Pajdla. Structure from many perspective images with occlusions. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, pages 355–369, 2002.
- [29] D. Martinec and T. Pajdla. Structure from many perspective images with occlusions. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, pages 355–369, May 2002.
- [30] S.J. Maybank and O.D. Faugeras. A theory of self calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

- 
- [31] C. McGlone. Bundle adjustment with object space geometric constraints for site modeling. In D.M. McKeown and Dowman, editors, *Proceedings of the SPIE Conference on Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II, Orlando, Florida, USA*, volume 2486, pages 25–36, April 1995.
- [32] P. McLauchlan, X. Shen, A. Manassis, P. Palmer, and A. Hilton. Surface-based structure-from-motion using feature groupings. In *Proceedings of the Fourth Asian Conference on Computer Vision*, 2000.
- [33] M. Pollefeys and L. Van Gool. A stratified approach to metric self-calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 407–412. IEEE Computer Society Press, June 1997.
- [34] M. Pollefeys, L. Van Gool, and M. Proesmans. Euclidean 3D reconstruction from image sequences with variable focal lengths. In *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1064, pages 31–42. Springer-Verlag, April 1996.
- [35] P. Poulin, M. Ouimet, and M.C. Frasson. Interactively modeling with photogrammetry. In *Proceedings of Eurographics Workshop on Rendering 98*, pages 93–104, June 1998.
- [36] I.D. Reid and D.W. Murray. Active tracking of foveated feature clusters using affine structure. *International Journal of Computer Vision*, 18(1):41–60, April 1996.
- [37] C. Rother and S. Carlsson. Linear multi view reconstruction and camera recovery using a reference plane. *International Journal of Computer Vision*, 9(49(2/3)):117–141, 2002.
- [38] C. Rother, S. Carlsson, and D. Tell. Projective factorization of planes and cameras in multiple views. In *Proceedings of the 15th International Conference on Pattern Recognition, Quebec, Canada*, pages 737–740, 2002.
- [39] G. Sparr. Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, volume I, pages 328–333. IEEE Computer Society Press, August 1996.
- [40] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 1100–1105, June 1997.
- [41] P. Sturm. Algorithms for plane-based pose estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pages 1010–1017, June 2000.
- [42] P. Sturm and S. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pages 432–437, June 1999.

- [43] P. Sturm and S.J. Maybank. A method for interactive 3D reconstruction of piecewise planar objects from single images. In *Proceedings of the tenth British Machine Vision Conference, Nottingham, England*, pages 265–274, September 1999.
- [44] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065, pages 709–720, April 1996.
- [45] R. Szeliski and P.H.S. Torr. Geometrically constrained structure from motion : Points on planes. In *3D Structure from Multiple Images of Large-scale Environments SMILE'98*. Springer Verlag, June 1998.
- [46] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.
- [47] B. Triggs. Factorization methods for projective structure and motion. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 845–851, 1996.
- [48] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 609–614. IEEE Computer Society Press, June 1997.
- [49] B. Triggs. Autocalibration from planar scenes. In *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, 1998.
- [50] R.Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Miami Beach, Florida, USA*, pages 364–374, 1986.
- [51] F. A. van den Heuvel and G. Vosselman. Efficient 3D-modeling of buildings using a priori geometric object information. In *Proceedings of the SPIE Conference on Videometrics V*, volume 3174, pages 38–49, 1997.
- [52] T. Werner and A. Zisserman. New techniques for automated architectural reconstruction from photographs. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, pages 541–555, May 2002.
- [53] M. Wilczkowiak, E. Boyer, and P. Sturm. Camera calibration and 3D reconstruction from single images using parallelepipeds. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, volume 1, pages 142–148, July 2001.
- [54] M. Wilczkowiak, E. Boyer, and P. Sturm. 3D modelling using geometric constraints: A parallelepiped based approach. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, volume 2353, pages 221–236, May 2002.

- [55] M. Wilczkowiak, P. Sturm, and E. Boyer. The analysis of ambiguous solutions in linear systems and its application to computer vision. In *To Appear in Proceedings of the 14th British Machine Vision Conference, Norwich, England*, September 2003.
- [56] L. Zelnik-Manor and M. Irani. Multi-view constraints on homographies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):214–223, February 2002.
- [57] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, September 1999.
- [58] A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. *Philosophical Transactions of the Royal Society of London, SERIES A*, 356(1740):1193–1211, 1998.



---

Unité de recherche INRIA Rhône-Alpes  
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes  
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)

<http://www.inria.fr>

ISSN 0249-6399