



**HAL**  
open science

# Camera Calibration and 3D Reconstruction from Single Images Using Parallelepipeds

Marta Wilczkowiak, Edmond Boyer, Peter Sturm

► **To cite this version:**

Marta Wilczkowiak, Edmond Boyer, Peter Sturm. Camera Calibration and 3D Reconstruction from Single Images Using Parallelepipeds. 8th International Conference on Computer Vision (ICCV '01), Jul 2001, Vancouver, Canada. pp.142-148, 10.1109/ICCV.2001.937510 . inria-00525657

**HAL Id: inria-00525657**

**<https://hal.inria.fr/inria-00525657>**

Submitted on 26 May 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Camera Calibration and 3D Reconstruction from Single Images Using Parallelepipeds

Marta Wilczkowiak\* Edmond Boyer Peter Sturm

Movi–Gravir–Inria Rhône-Alpes, 655 Avenue de l’Europe, 38330 Montbonnot, France

Firstname.Surname@inrialpes.fr

## Abstract

*In this paper, parallelepipeds and their use in camera calibration and 3D reconstruction processes are studied. Parallelepipeds naturally characterize rigidity constraints present in a scene, such as parallelism and orthogonality. A subclass of parallelepipeds – the cuboids – has been frequently used over the past to partially calibrate cameras. However, the full potential of parallelepipeds, in camera calibration as well as in scene reconstruction, has never been clearly established. We propose a new framework for the use of parallelepipeds which is based on an extensive study of this potential. In particular, we exhibit the complete duality that exists between the intrinsic metric characteristics of a parallelepiped and the intrinsic parameters of a camera. Our framework allows to fully exploit parallelepipeds and thus overcomes several limitations of calibration approaches based on cuboids. To illustrate this framework, we present an original and very efficient interactive method for 3D reconstruction from single images. This method allows to quickly build a scene model from a single uncalibrated image.*

## 1. Introduction

In this paper we describe how to use rigidity constraints through parallelepipeds to infer information on a camera, as well as on the environment, from images. The motivation is to propose straightforward and practical methods for camera calibration and scene modeling with few interactions and little required *a priori* knowledge. The interest arises in 3D modeling applications using images with no information on the cameras being used, for example applications using single images obtained from the web.

Rigidity constraints have been widely studied in the computer vision community over the last decade. In a seminal work, Caprile and Torre [2] used cuboids, i.e. parallelepipeds with right angles, to estimate some camera parameters. Their approach is based on vanishing points defined by the cuboid’s projected edges. Such vanishing points correspond to perpendicular directions in space and

impose therefore constraints on the transformation between 3D space and the image plane. Following this idea, several approaches, which make use of vanishing points and lines, have been proposed to either partially calibrate cameras or reconstruct scenes [8, 4, 10, 5, 12, 7]. However, computing vanishing points and lines in the image is often numerically unstable. Moreover, dealing with individual vanishing points does not allow to fully exploit the redundancy contained in the input, i.e. that all the vanishing points stem from the projection of a single parallelepiped. In contrast to the above mentioned approaches, we do not compute vanishing points or lines but projection matrices such that parallelepiped’s projections fit the concerned image points.

Recently, calibration results have also been obtained using planar patterns and homographies [11, 13, 14, 15]. While more flexible than standard calibration techniques, homography based approaches require either Euclidean information or, for self-calibration, many images in general position. In addition, planar structures used for calibration are often part of parallelepipedic structures which are not fully exploited, much as the vanishing point/line methods cited above. In this sense, our approach is a generalization of plane-based methods with metric information to three-dimensional, parallelepipedic patterns.

Other works partially use parallelepipeds. For example, in [6] cuboids are used for 3D reconstruction, but several calibrated images are required. Also, in [3] parallelepipeds are used for calibration in augmented reality applications. However, the proposed approach has a limited application domain since: the camera must satisfy a strong constraint – unit aspect ratio – and only partial knowledge on the parallelepiped – angles – can be used for calibration.

Our approach is based on the duality between the intrinsic characteristics of a camera and those of a parallelepiped. Given an image of a parallelepiped, each known intrinsic parameter of either camera or parallelepiped gives a constraint on the parameter sets of both entities. Since parallelepipeds are frequently present in man-made environments (roofs, facades, etc), a rich variety of images can be calibrated using this approach. Furthermore, little knowledge on parallelepipeds is generally required. For exam-

---

\*This work was supported by the EU project IST-1999-10756, VISIRE.

ple, two right angles associated to provably valid assumptions on camera intrinsics are sufficient for calibration. This makes parallelepipeds a very flexible tool which is well adapted to unknown environments. Moreover, parallelepipedic structures are a natural building block for realistic models of man-made scenes which can be easily extended using simple subsequent constraints, such as coplanarity of points etc. We present a method for interactive scene modeling from single images, using these constraints.

The paper is organized as follows. Section 2 introduces definitions that are used in the rest of the paper. Calibration using parallelepipeds is studied in Section 3. The duality between intrinsic camera and parallelepiped parameters is established. An exhaustive study of singular configurations for calibration is given and experimental results are provided to precisely define the application context. The single image based 3D reconstruction approach is presented in Section 4 and illustrated by realistic scene models.

## 2. Definitions

**Camera Model** We assume that the imaging system is based on the pinhole model (i.e., perspective projection). The projection from 3D space to the 2D image plane can thus be expressed by:  $p \sim M \cdot P$ , where  $M$  is a 3x4 matrix and  $P, p$  are the homogeneous point coordinates in 3D space and in the image plane respectively. The matrix  $M$  can be decomposed as:  $M = K \cdot [R \ t]$ .  $[R \ t]$  is the 3x4 matrix determining the relative orientation  $R$  and position  $-R^T t$  of the camera in 3D space and  $K$  is the 3x3 calibration matrix defining the pinhole camera:

$$K = \begin{bmatrix} \alpha_u & s = 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix},$$

where  $\alpha_u$  and  $\alpha_v$  stand for the focal length, expressed in horizontal and vertical pixel dimensions,  $s$  is a skew parameter considered as equal to zero (but this can be relaxed) and  $u_0, v_0$  are the pixel coordinates of the intersection of the optical axis with the image plane.

**Parallelepipeds** A parallelepiped is defined by twelve parameters: six extrinsic parameters describing its orientation and position in 3D space (three position and three orientation parameters) and six intrinsic parameters that describe its Euclidean shape: three dimension parameters (edge lengths) and the three angles between parallelepiped edges (see figure 2). The perspective projection of a parallelepiped depends only on eleven parameters due to the projection's scale-depth ambiguity.

Thus, five image points and an image direction are sufficient to completely define the projection of a parallelepiped. Other parallelepiped's image points can then be computed via classical three point perspective, using the parallelism of parallelepiped edges. Hence, parallelism in 3D does not

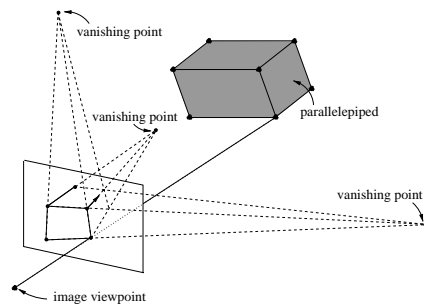


Figure 1: A parallelepiped and its projection which is completely defined by five image points and an image direction. The parallelepiped's projected edges define three vanishing points in the image plane.

provide any information to compute the 3D to 2D projection, and thus for calibration (what *is* useful for calibration is knowledge of a parallelepiped's *intrinsic* parameters, i.e. angles between non parallel edges and length ratios). Equivalently, additional image points, besides the minimum of five and a half, do not provide calibration constraints. However, they add useful redundancy for a more accurate computation of the projection and eventually, for calibration.

## 3. Camera calibration

Classical ways to derive calibration constraints from structures containing parallelism consist in first determining vanishing points and then expressing the fact that the corresponding space directions are, e.g. orthogonal [2, 7], which leads to equations on the intrinsic parameters. Vanishing points are usually estimated as intersections of lines, and more precisely in real situations, as points which minimize distances to projections of parallel lines [2, 9, 10]. Such approaches often suffer from numerical instability, especially in cases of near parallelism of image edges. Furthermore, in the case of parallelepipeds, these approaches do not take into account the fact the concerned 2D points belong to a parallelepiped projection (as discussed in the introduction).

In contrast to this, we do not compute explicitly vanishing points but we first estimate the best projection matrix such that the defined image points correspond to a parallelepiped in space. The error between reprojected and measured image points is minimized, thus avoiding the intersection of (multiple) lines, and all information is used simultaneously. From this projection matrix and *a priori* knowledge on the parallelepiped (angles, length ratios), it is easy to derive constraints on the intrinsic parameters of the used camera, most of these constraints being linear. These constraints are then used to estimate intrinsic parameters. They can be combined in a linear framework with assumptions or *a priori* knowledge on the camera's intrinsic parameters.

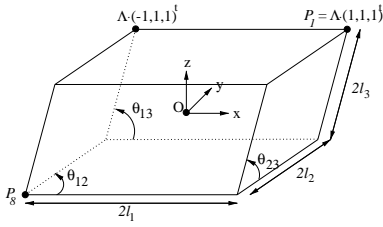


Figure 2: Parameterization of parallelepipeds:  $2l_i$  are the edge lengths;  $\theta_{ij}$  are the angles between non parallel edges.

### 3.1. Parallelepiped projection matrix

As stated above, a parallelepiped is defined by three orientation parameters, three position parameters, three dimension parameters and three angles. Let  $R$ ,  $-R^T t$ ,  $(l_1, l_2, l_3)$  and  $(\theta_{12}, \theta_{23}, \theta_{13})$  be these parameters respectively. Then, image projections  $p_{i \in [1..8]} = [u_i \ v_i]$  of the parallelepiped vertices satisfy:

$$\begin{pmatrix} \alpha_1 u_1 & \dots & \alpha_8 u_8 \\ \alpha_1 v_1 & \dots & \alpha_8 v_8 \\ \alpha_1 & \dots & \alpha_8 \end{pmatrix} = \tilde{X} \cdot \begin{pmatrix} 1 & \dots & -1 \\ 1 & \dots & -1 \\ 1 & \dots & -1 \\ 1 & \dots & 1 \end{pmatrix}. \quad (1)$$

$\tilde{X}$  is the 3x4 projection matrix which is defined up to a scalar factor by:

$$\tilde{X} \sim M \cdot \Lambda \sim K \cdot [R t] \cdot \tilde{\Lambda}, \quad (2)$$

where the matrix  $\tilde{\Lambda}$  depend on the parameterization that is chosen for parallelepipeds. We chose <sup>1</sup> (see Fig. 2):

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

with:  $c_{ij} = \cos \theta_{ij}$ ,  $s_{ij} = \sin \theta_{ij}$ ,  $\theta_{ij} \in ]0 \ \pi[$ ,  $l_i > 0$ . Note that in the case of cuboids, the matrix  $\tilde{\Lambda}$  is diagonal.

If at least six vertices  $p_i$  of the parallelepiped's projection are known (or five vertices and a direction e.g.), then  $\tilde{X}$  can be estimated, up to a scalar factor, from equation (1) by a linear least squares method.  $\tilde{X}$  results from the composition of the projection matrix  $M$  and an affinity  $\tilde{\Lambda}$ , thus, in the general case, no constraints such as orthonormality apply to its lines or columns. Consequently, parameters of the matrix  $\tilde{X}$  can be simply estimated by solving the homogeneous linear system of equations obtained from equation (1) and the available image coordinates  $[u_i \ v_i]$  (we use SVD to this purpose).

### 3.2. Intrinsic parameter duality

The projection matrix  $\tilde{X}$  defined in the previous section captures all the geometric information given by a parallelepiped's projection. Now consider the 3x3 matrix

<sup>1</sup>Different parameterizations can be used. For illustration, we chose one with a simple physical interpretation.

$X \sim K \cdot R \cdot \Lambda$ , where  $X$  and  $\Lambda$  are the matrices of the first three lines and columns of  $\tilde{X}$  and  $\tilde{\Lambda}$  respectively. From equation (1), we get:

$$X^T \cdot K^{-T} \cdot K^{-1} \cdot X \sim \Lambda^T \cdot \Lambda, \quad (3)$$

where  $K^{-T} \cdot K^{-1} = \omega$  is the matrix of the absolute conic's image (IAC) which depends on the intrinsic camera parameters, and  $\Lambda^T \cdot \Lambda = \mu$  is an analogous entity, depending on the parallelepiped's intrinsic parameters:

$$\mu = \Lambda^T \cdot \Lambda = \begin{pmatrix} l_1^2 & l_1 l_2 \cos \theta_{12} & l_1 l_3 \cos \theta_{13} \\ l_1 l_2 \cos \theta_{12} & l_2^2 & l_2 l_3 \cos \theta_{23} \\ l_1 l_3 \cos \theta_{13} & l_2 l_3 \cos \theta_{23} & l_3^2 \end{pmatrix}.$$

Equation (3) shows the duality that exists between a camera and a parallelepiped. They both have intrinsic parameters:  $K$  for a camera and  $\Lambda$  for a parallelepiped which are dual by equation (3). Thus, knowledge on the intrinsic parameters of either the parallelepiped or the camera leads to constraints on the parameters of the respective other entity. For example, the assumption of the camera's skew parameter being zero gives a constraint on the parallelepiped's intrinsic parameters. In the following, we detail typical constraints on the intrinsic camera parameters.

Let  $X_i$  be the  $i$ th column of  $X$ , then the duality equation (3) leads to (non-exhaustive list):

1. a known angle  $\theta_{ij}$  gives the following quadratic constraint on the elements of  $\omega$ :
$$(X_i^T \cdot \omega \cdot X_j)^2 - c_{ij}^2 (X_i^T \cdot \omega \cdot X_i) (X_j^T \cdot \omega \cdot X_j) = 0;$$
2. a known *right* angle  $\theta_{ij}$  gives the following linear constraint:  $X_i^T \cdot \omega \cdot X_j = 0$ ;
3. a known length ratio  $r_{ij} = l_i/l_j$  gives the following linear constraint:  $X_i^T \cdot \omega \cdot X_i - r_{ij}^2 X_j^T \cdot \omega \cdot X_j = 0$ .

A few remarks are in order:

- The columns  $X_i$  of  $X$  are the homogeneous coordinates of the vanishing points associated to the parallelepiped's projection.
- When several of the above constraints hold, then all deduced equations can be stacked together to solve for  $\omega$ . This is especially attractive for the linear constraints.
- Knowing the length  $l_i$  of an edge (instead of only a length ratio) does not provide any constraint on the intrinsic camera parameters, it simply cancels the depth-scale ambiguity of the projection.

In the rest of the paper, we focus on linear constraints which lead to a direct linear estimation of the intrinsic camera parameters. Quadratic constraints can be used subsequently to the linear ones, to obtain a finite number of solutions in case the linear system is underconstrained.

### 3.3. Estimating intrinsic camera parameters

Suppose that four or more linear constraints as described in the previous section are available from one or several parallelepipeds, then the elements  $\omega_{ij}$  of  $\omega$  can be estimated up to a scalar factor by a linear least squares approach. This is achieved by solving a homogeneous linear equation system of the type  $Ax = 0$ , where  $x$  is the vector of the elements  $\omega_{ij}$  not generally equal to zero. From the elements  $\omega_{ij}$ , the intrinsic parameters are simply extracted as<sup>2</sup>:

$$\begin{aligned} u_0 &= -\frac{\omega_{13}}{\omega_{11}}, & v_0 &= -\frac{\omega_{23}}{\omega_{22}}, \\ \alpha_u^2 &= \frac{\omega_{11}\omega_{22}\omega_{33} - \omega_{11}\omega_{23}^2 - \omega_{22}\omega_{13}^2}{\omega_{11}^2\omega_{22}}, & \alpha_v^2 &= \frac{\omega_{11}}{\omega_{22}}\alpha_u^2. \end{aligned} \quad (4)$$

**Using prior knowledge or assumptions** If less than four constraints are available, then the camera can not be fully calibrated. Still, if some of the intrinsic parameters are known (or assumed to have some standard value), then the other parameters can in general be calibrated if their number is less or equal to that of the available constraints. For instance, the aspect ratio  $\alpha_u/\alpha_v$  may be given by the camera manufacturer, or, as successfully experimented in this paper (see Sec. 3.5), the assumption that the optical axis intersects the image plane at the image center can be made (i.e.,  $u_0$  and  $v_0$  correspond to half the image sizes). In order to well use such information, and thus to separate in the estimation process known parameters from unknown parameters, we decompose the matrix  $K$  into two matrices:  $K = K_k \cdot K_u$ , where  $K_k$  is the matrix of known, or assumed known parameters, and  $K_u$  the matrix of unknown parameters. The image projection  $p_i$  of a parallelepiped vertex can then be transformed accordingly to:  $p'_i = K_k^{-1} \cdot p_i$ . The parallelepiped projection matrix  $\tilde{X}'$  associated with the  $p'_i$ 's is then:

$$\tilde{X}' \sim K_k^{-1} \cdot K \cdot [R \ t] \cdot \tilde{\Lambda} \sim K_u \cdot [R \ t] \cdot \tilde{\Lambda}.$$

This matrix can be estimated from the image points  $p'_i$  by the same linear approach as for  $\tilde{X}$ . This leads to simplified forms for the matrix  $\omega = K_u^{-T} \cdot K_u^{-1}$ .

However, care should be taken when using prior knowledge since it can lead to bad conditioning of the linear equation systems which are solved. Note also that an alternative approach for using known intrinsic camera parameters is to include additional equations entering the linear least squares solution. This might be desirable in cases of only approximate prior knowledge, where weighted least squares can be used to adjust that knowledge.

### 3.4. 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

<sup>2</sup>With five linear constraints, we could similarly solve for the skew parameter.

to a singularity, the results become usually very unstable. In this section, we examine the singularities for the linear calibration algorithm described above. We have studied all possible combinations of *a priori* knowledge, on both camera and parallelepiped intrinsic parameters. Due to lack of space, we only sketch the methodology followed.

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 (3) 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 intrinsics. Let  $K = K_k \cdot K_u$  be the true calibration matrix, and  $K' = K_k \cdot K'_u$  the estimated one (we decompose in known and unknown parts, as in the previous section, so  $K'$  and  $K$  share of course the known part  $K_k$ ). The same holds for the parallelepiped:  $\Lambda = \Lambda_k \cdot \Lambda_u$  and  $\Lambda' = \Lambda_k \cdot \Lambda'_u$  are respectively the true and estimated intrinsic parameters. A singularity exists if there are solutions with  $K'_u \neq K_u$  and  $\Lambda'_u \neq \Lambda_u$ . From equation (3), it is easy to derive the following equality, which is defined up to scale (using  $X \sim K_k \cdot K_u \cdot R \cdot \Lambda_u \cdot \Lambda_k$ ):

$$R^T \cdot K_u^{-T} \cdot K'_u^{-T} \cdot K'_u^{-1} \cdot K_u \cdot R = \Lambda_u^{-T} \cdot \Lambda'^T_u \cdot \Lambda'_u \cdot \Lambda_u^{-1}.$$

A singularity, as defined above, is then equivalent to the existence of matrices  $\omega'' = K_u^T \cdot K'_u^{-T} \cdot K'_u^{-1} \cdot K_u$  and  $\mu'' = \Lambda_u^{-T} \cdot \Lambda'^T_u \cdot \Lambda'_u \cdot \Lambda_u^{-1}$  that are different from the identity. Depending on the *a priori* knowledge,  $\omega''$  and  $\mu''$  have special forms (as shown in tables 1 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  $\mu''$  different from the identity, and of the required special form, to:

$$R^T \cdot \omega'' \cdot R \sim \mu''$$

Based on this definition, it is a rather mechanical, though sometimes tricky, task, to derive singular relative orientations. Table 2 shows all singularities for parallelepipeds with two or three right angles (other cases are not shown due to lack of space). We describe the singularities in geometrical terms, by describing the relative orientation of the parallelepiped with respect to the camera.

Known camera intrinsics			
(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 1: Structure of  $\omega''$  depending on prior knowledge on intrinsic camera parameters. Structure of  $\mu''$  is similar.

**Three right angles, two length ratios.** In this case, the Euclidean structure of the parallelepiped is completely

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

Table 2: Singular relative orientations for various combinations of prior knowledge on intrinsic parameters. Cases are denoted X-Y-Z, where X refers to table 1 and Y and Z are the number of known right angles respectively length ratios. For further explanations, see text.

given (up to scale), and it can be used as a classical calibration object. There are no general singularities proper to the use of a parallelepiped, but of course the generic singularities described in [1] apply here too.

**Three right angles, one length ratio (cases \*-3-1 in table 2).** In table 2,  $v$  represents any of the four parallelepiped’s edges with unknown length (ratio).

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

### 3.5. Experimental results

Synthetic data were used to evaluate the sensitivity of the calibration algorithm in the presence of noise. The concerned noisy data are the vertices’ projections of the reference parallelepiped and the *a priori* known camera parameters (e.g. principal point). Calibration tests were performed for various relative orientations of the parallelepiped and the camera, which allowed us to experimentally confirm singularities. The parallelepiped’s parameters were:  $\{\theta_{12}, \theta_{13}, \theta_{23}\} = \{\pi/2, \pi/3, \pi/2\}$  and  $\{l_1, l_2, l_3\} =$

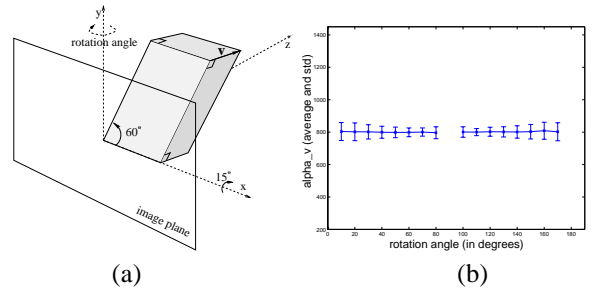


Figure 3: (a) illustration of experiment; (b)  $\alpha_v$  (average and standard deviation, true value is 800) as a function of the relative orientation between the parallelepiped and the camera, only two right angles are known.

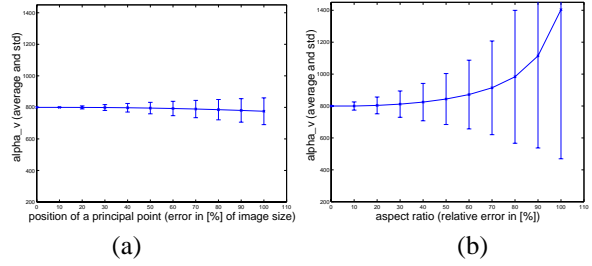


Figure 4:  $\alpha_v$  (average and standard deviation, true value is 800) as a function of the difference between assumed and true values of: (a) the principal point; (b) the aspect ratio.

$\{120, 250, 130\}$ . Images were generated with the following camera intrinsics:  $\alpha_u = 500, \alpha_v = 800, u_0 = v_0 = 256$  and for a  $512 \times 512$  image. Generally, only information about right angles was used in the calibration process. For tested scenarios, 300 experiments with various noise amplitudes were performed. Tables given below show the average and standard deviation of the calibration results for  $\alpha_v$ . Experiments on the sensitivity of the calibration to noise in image point positions reveal that the error for  $\alpha_v$  grows linearly with the noise amplitude. Figure 3 shows calibration errors for a two pixel uniform noise and various orientations of the parallelepiped. In this experiment, a parallelepiped with two right angles, rotated  $15^\circ$  around a horizontal axis parallel to the image plane, was rotated around a vertical axis parallel to the image plane (Fig. 3-(a)). Figure 3-(b) shows that singularities occur when the axis between two rectangular faces of the parallelepiped is parallel to the image plane or orthogonal to the image  $x$ -axis (rotation angles of  $90^\circ$  and of  $0^\circ/180^\circ$ ). This corresponds to case C-2-0 in table 2. Figure 4 shows the influence on calibration of wrong assumptions concerning the principal point position and the aspect ratio. In figure 4-(a), we can see that the error on  $\alpha_v$  grows nearly linearly with the distance between the assumed principal point and the true one. Even if the principal point is assumed to lie on the image border,

the relative error is still less than 11%. This confirms that it is reasonable to assume that the principal point is at the image's center. On the other hand, figure 4-(b) shows that  $\alpha_v$ 's estimation degrades quickly (quadratically) when the difference between the aspect ratio's assumed and true values increases. This implies that assumptions on the aspect ratio are feasible only if the assumed value is rather close to the true one.

## 4. Single image based reconstruction

An important application to parallelepiped based calibration is interactive 3D reconstruction from single images. This presents a major advantage over other methods: simplicity. Indeed only one image is required and just a small amount of effort is needed for both calibration and reconstruction: a few points must be picked in the image to define the primitives' image positions. It thus appears 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.

The reconstruction approach that is presented in this section is independent from the calibration process previously presented. The idea is to use, for reconstruction, a set of points such that all points in the set satisfy constraints common to subsets of them. Such constraints are for example coplanarity, collinearity, or the fact that points belong to a parallelogram. This allows the reconstruction of a connected polygonal structure which includes the parallelepipeds<sup>3</sup> that are used for calibration, as well as points subsequently defined in the image. The final model is thus composed of connected structures, each of them corresponding to one or several linear constraints. The reconstruction is obtained by solving a single linear system which regroups all the defined constraints.

Typical linear constraints on scene points are:

1. Parallelogram: as noted in [3], four points  $P_{i \in [1..4]}$  in space which belong to a parallelogram satisfy the following linear relation (assuming they are ordered):  $P_1 - P_2 + P_3 - P_4 = 0$ . Viewing lines for points can easily be determined when intrinsic parameters are known:  $K^{-1} \cdot p_i$  is parallel to the viewing direction  $d_i$  of the point  $P_i$  with image  $p_i$ . Thus the above equation can be transformed into:  $\alpha_1 d_1 - \alpha_2 d_2 + \alpha_3 d_3 - \alpha_4 d_4 = 0$ , where the  $\alpha_i$ 's are the unknown *depths* of points  $P_i$ . This equation can be written for a set of connected parallelograms in the scene. The resulting linear system can be solved up to a scale factor (by fixing, for instance, one depth  $\alpha_i$ ). However, for unconnected sets of parallelograms, further constraints (coplanarity,

<sup>3</sup>For parallelepipeds, individual reconstructions can also be achieved using equation (2).

collinearity, etc) must be imposed to obtain a consistent global reconstruction.

2. Coplanarity: four coplanar points  $P_{i \in [1..4]}$  satisfy:  $P_4 - P_1 + \beta_1(P_3 - P_1) + \beta_2(P_2 - P_1) = 0$ . Hence, the depths  $\alpha_i$  of the  $P_i$  (i.e.  $\alpha_i d_i = P_i$ ) satisfy:

$$(-\beta_1 - \beta_2 - 1)\alpha_1 d_1 + \beta_2 \alpha_2 d_2 + \beta_1 \alpha_3 d_3 + \alpha_4 d_4 = 0,$$

which is linear in  $(-\beta_1 - \beta_2 - 1)\alpha_1, \beta_2 \alpha_2, \beta_1 \alpha_3$  and  $\alpha_4$ . This equation can be used to solve for one point's depth or for the unknown relative scale between two reconstructed point sets. A similar argument holds for collinear points.

The complete calibration and reconstruction algorithm consists of the following stages:

1. The projections of one or more parallelepipeds are defined in the image and used to determine the intrinsic camera parameters.
2. Other image points and their associated linear dependencies (parallelograms, etc.) are defined, forming a linear system. It is solved for relative depths of all points. This includes parallelepiped vertices for which the above parallelogram dependencies hold.
3. A 3D textured model is computed using reconstructed points and the input image.

### 4.1. Experimental results

We have developed an interface based on OpenGL to implement the ideas presented in this paper. The motivation is to have an easy-to-use tool allowing models to be computed in a few minutes at most. Models can also be exported in the standard VRML format. We present screenshots of two reconstructed models which illustrate the potential of the method.

**Reconstruction of an indoor scene** Figure 5 shows the original image and screenshots of the reconstruction of an indoor scene. Most of the angles in the original scene are not perfectly right (average deviation of approximately 3°). Calibration was based on the cupboard seen in the central part of the image. For reconstruction different types of constraints were used - 3 sets of parallelograms (main part of the image, window and wooden belt) were connected using coplanarity constraints. Walls were reconstructed using coplanarity and collinearity constraints. The final 3D model was evaluated by measuring angles that were assumed to be right. The maximum relative error for these angles is 1.8%.

**Reconstruction of an outdoor scene** The original image, taken with another camera, showed moderate optical distortion, which was corrected using reference distortion parameters (Fig. 6 shows the undistorted image and also the reference primitives used for calibration). The aspect ratio was calibrated with 1.8% of relative error and  $\alpha_v$  with 1.9%. The angles of the tower, in the 3D reconstruction, are all within 5% of being right.



Figure 5: Original image and screenshots of the reconstruction of the indoor scene.



Figure 6: Original image and screenshots of the reconstruction of the outdoor scene.

## 5 Conclusion

We have presented a framework for camera calibration and 3D scene reconstruction based on parallelepipeds. We have shown throughout the paper that: (i) parallelepipeds and perspective cameras are dual; (ii) knowledge of right angles and length ratios, as well as prior knowledge on intrinsic camera parameters can be used simultaneously in a linear framework; (iii) assumptions on the principal point's position are robust and more robust than those on the aspect ratio; (iv) accurate calibration can be achieved using one or more of the widely present parallelepipeds that have two right angles. Singularities of the calibration approach have been clearly identified and are given for the most general situations (two or three right angles). They delimit the context within which parallelepipeds should be used. The proposed framework was illustrated by single image based reconstruction examples. The reconstructed models demonstrate the potential of the method for a general public use.

## References

- [1] T. Buchanan, "The Twisted Cubic and Camera Calibration," *CVGIP*, 42(1), pp. 130–132, 1988.
- [2] B. Caprile and V. Torre, "Using Vanishing Points for Camera Calibration," *IJCV*, 4(2), pp. 127–140, 1990.
- [3] C.S. Chen, C.k. Yu, and Y.P. Hung, "New Calibration-free Approach for Augmented Reality Based on Parameterized Cuboid Structure," *ICCV*, pp. 30–37, 1999.
- [4] R. Cipolla and E. Boyer, "3D model acquisition from uncalibrated images," *IAPR Workshop on Machine Vision Applications*, pp. 559–568, 1998.
- [5] A. Criminisi, I. Reid, and A. Zisserman, "Single View Metrology," *ICCV*, pp. 434–441, 1999.
- [6] P.E. Debevec, C.J. Taylor, and J. Malik, "Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach," *SIGGRAPH*, 1996.
- [7] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.
- [8] Y. Horry, K. Anjyo, and K. Arai, "Tour Into the Picture," *SIGGRAPH*, 1997.
- [9] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Lecture Note, 1995.
- [10] D. Liebowitz and A. Zisserman, "Metric Rectification for Perspective Images of Planes," *CVPR*, pp. 482–488, 1998.
- [11] E. Malis and R. Cipolla, "Multi-View Constraints Between Collineations: Application to Self-Calibration from Unknown Planar Structures," *ECCV*, pp. 610–624, 2000.
- [12] P.F. Sturm and S.J. Maybank, "A Method for Interactive 3D Reconstruction of Piecewise Planar Objects from Single Images," *BMVC*, pp. 265–274, 1999.
- [13] P.F. Sturm and S.J. Maybank, "On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications," *CVPR*, pp. 432–437, 1999.
- [14] B. Triggs, "Autocalibration from Planar Scenes," *ECCV*, pp. 89–105, 1998.
- [15] Z. Zhang, "A flexible new technique for camera calibration," *IEEE-PAMI*, 22(11), pp. 1330–1334, 2000.