

Camera calibration and 3D reconstruction from a single view based on scene constraints

Guanghai Wang^{a,b,*}, Hung-Tat Tsui^a, Zhanyi Hu^b, Fuchao Wu^b

^aDepartment of Electronic Engineering, The Chinese University of Hong Kong, Shatin, N.T. Hong Kong, P. R. China

^bNational Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P. R. China

Received 30 January 2004; received in revised form 27 July 2004; accepted 29 July 2004

Abstract

This paper mainly focuses on the problem of camera calibration and 3D reconstruction from a single view of structured scene. It is well known that three constraints on the intrinsic parameters of a camera can be obtained from the vanishing points of three mutually orthogonal directions. However, there usually exist one or several pairs of line segments, which are mutually orthogonal and lie in the pencil of planes defined by two of the vanishing directions in the structured scenes. It is proved in this paper that a new independent constraint to the image of the absolute conic can be obtained if the pair of line segments is of equal length or with known length ratio in space. The constraint is further studied both in terms of the vanishing points and the images of circular points. Hence, four independent constraints on a camera are obtained from one image, and the camera can be calibrated under the widely accepted assumption of zero-skew. This paper also presents a simple method for the recovery of camera extrinsic parameters and projection matrix with respect to a given world coordinate system. Furthermore, several methods are presented to estimate the positions and poses of space planar surfaces from the recovered projection matrix and scene constraints. Thus, a scene structure can be reconstructed by combining the planar patches. Extensive experiments on simulated data and real images, as well as a comparative test with other methods in the literature, validate our proposed methods.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Camera calibration; 3D reconstruction; Absolute conic; Circular points; Single view modeling

1. Introduction

3D reconstruction from 2D images is a central problem of computer vision. Examples and applications of this task include robot navigation and obstacle recognition, augmented reality, architectural surveying, forensic science and others. The classical method for this problem is to reconstruct the metric structure of the scene from two or more images by stereovision techniques [1,2]. However, this is a hard task due to the problem of seeking correspondences between different views. In recent years, some attentions are focused on reconstruction directly from a single uncalibrated image. It is well known that only one image cannot provide enough information for a complete 3D reconstruction.

However, some metrical quantities can be inferred directly from a single image with the prior knowledge of geometrical scene constraints. Such constraints may be expressed in terms of vanishing points or lines, co-planarity, special inter-relationship of features and camera constraints.

There are many studies on the problem of single view based calibration and reconstruction in the literature. Traditional approaches for solving this problem utilize a particular cue, such as shading, lighting, texture and defocusing [22,23]. These methods make strong assumptions on shape, reflectance or exposure, and tend to require a controlled environment, which is often not practical. The popular approaches in recent years are trying to use the geometrical information obtained from images. Horry et al. [3] propose a technique, named *tour into the picture*. They created a graphic user interface that allows the user to separate a 2D image into background and foreground, and separate the background into five regions and form a cube. The foreground images are then placed inside the cube at

* Corresponding author. Address: Aviation University of Airforce, Changchun, 130022, P.R. China. Tel.: +86 431 6958527; fax: +86 431 6958671.

E-mail address: ghwang@ee.cuhk.edu.hk (G. Wang).

appropriate locations, and a virtual ‘walk through’ animation is generated finally. Zhang et al. [4] propose a method, which combines a sparse set of user-specified constraints, such as surface position, normals, silhouettes and creases, to generate a well-behaved 3D surface satisfying these constraints.

For a wide variety of man-made environment (architecture, facades [5], etc.), a cuboid is a reasonable model. Caprile and Torre [6] propose a method for camera calibration from vanishing points computed from the projected edges of the cuboids. These vanishing points correspond to three mutually orthogonal directions in space, which can provide three independent constraints to the intrinsic parameters of a camera. Following this idea, several approaches that make use of vanishing points and lines have been proposed for either cameras calibration or scene reconstruction [7–10]. Most of the studies are usually under the assumption of square pixels, i.e. zero-skew and unit aspect ratio. However, the assumption may not be applicable to some off-the-shelf digital cameras. Wilczkowiak et al. [11,12] and Chen et al. [13] expand the idea to general parallelepiped structures, and use the constraints of parallelepipeds for camera calibration. Wilczkowiak et al. [12] also present a complete duality that exists between the intrinsic metric characteristics of a parallelepiped and the intrinsic parameters of a camera. Criminisi et al. [14] study the problem by computing 3D affine measurement from a single perspective image. The approach is based on the vanishing line of a reference plane and the vanishing point in vertical direction.

Inspired by the idea of cuboid, our work is targeted on man-made structures, such as architectures, which typically contain three orthogonal principal directions, and the corresponding vanishing points can be retrieved from the image of straight lines using maximum likelihood estimator [7]. Our research aims at making full use of scene constraints to obtain a more accurate and photorealistic model of a 3D object. It is assumed that there are one or several pairs of mutually orthogonal line segments, which lie in the pencil of planes defined by two of the vanishing directions in the scene and the pair of segments are of equal length or with known length ratio. This is not rare for most of man-made objects.

The main contribution of this paper is that we prove that the pair of line segments with equal length or known length ratio in the scene can provide an additional independent constraint to the image of the absolute conic. Three equivalent forms of the constraints are further studied both in terms of the orthogonal vanishing points and the image of circular points. We also present a simple approach to the recovery of camera pose and projection matrix with respect to a given world system. Thus, the object can be reconstructed by taking measurement on piecewise planar patches [11,14].

The remaining parts of this paper are organized as follows. In Section 2, some preliminaries on projection

matrix and the absolute conic are reviewed. Then, the calibration method is elaborated in detail in Section 3. The method to recover the extrinsic parameters and camera projection matrix is given in Section 4. In Section 5, the methods for measurement and 3D reconstruction are presented. The test results with simulated data and real images are presented in Sections 6 and 7, respectively. The conclusion of this paper is given in Section 8.

2. Notation and preliminaries

In order to facilitate our discussions in the subsequent sections, some preliminaries on camera projection matrix and the absolute conic are presented here. Readers can refer to Hartley [1] and Faugeras [2] for more detail. In this paper, the following notations are used. An image point is denoted by a bold lower case letter, e.g. \bar{x} , while its corresponding homogeneous vector is denoted by $\mathbf{x} = [\bar{x}^T, w]^T$, $(x)_i$ denotes the i th element of vector \mathbf{x} ; a matrix is denoted by a bold upper case letter, e.g. \mathbf{P} , \mathbf{P}_i denotes the i th column vector of matrix \mathbf{P} , while $P_{i,j}$ stands for the element in the i th row and j th column of \mathbf{P} . ‘ \approx ’ stands for equality up to scale. For two line segments s and s' in an image, ‘ $s \cong s'$ ’ stands for their corresponding segments in the Euclidean space are equal in length.

2.1. Camera projection matrix

Under perspective projection, a 3D point \mathbf{x} in space is projected to an image point \mathbf{m} via a 3×4 rank 3 projection matrix \mathbf{P} as

$$\lambda \mathbf{m} = \mathbf{P}\mathbf{x} = \mathbf{K}[\mathbf{R}, \mathbf{t}]\mathbf{x} \quad (1)$$

where λ is a non-zero scalar; \mathbf{R} and \mathbf{t} are the rotation matrix and translation vector from the world system to the camera system. \mathbf{K} is the camera calibration matrix in the form of

$$\mathbf{K} = \begin{bmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f_u & f_u \text{ctg } \theta & u_0 \\ 0 & r f_u & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where f_u, f_v represent the camera’s focal length corresponding to the u and v axes of camera coordinates; (u_0, v_0) is the coordinates of the camera’s principal point; $s = f_u \text{ctg } \theta$ refers to the skew factor, with θ the included angle of the u and v axes; $r = f_v/f_u$ is termed as the aspect ratio. For most CCD cameras, we can assume rectangle pixels, i.e. $\theta = 90^\circ$ or $s = 0$. Then, the camera becomes a simplified one with only four intrinsic parameters. For some high quality cameras, we may even assume square pixels, i.e. $s = 0$ and $r = 1$ ($f_u = f_v$), and the camera model is simplified to three parameters accordingly.

Lemma 1. *The first three columns of projection matrix \mathbf{P} are images of the vanishing points corresponding to the X, Y*

and Z axes of the world system, respectively, and the last column P_4 is the image of the origin of the world system.

Lemma 2. The plane at infinity in space can be expressed as $\Pi_\infty = [0, 0, 0, 1]^T$, and the mapping between Π_∞ and the image of the infinite plane π_∞ is the planar homography $H_\infty = KR$.

Lemma 3. Suppose the homography between a space plane and the image plane is H , then for a conic $C(x^T C x = 0)$ on the space plane, its corresponding image is: $C' = H^{-T} C H^{-1}$.

2.2. The absolute conic and the image of the absolute conic

The absolute conic (AC), Ω_∞ , is a conic on the plane at infinity Π_∞ , which satisfies

$$\Omega_\infty = \{\bar{x}_\infty | \bar{x}_\infty^T \bar{x}_\infty = 0\} \quad (3)$$

where \bar{x}_∞ is an infinite point on Π_∞ . Thus, Ω_∞ corresponds to a conic $C_\infty = I$. It is a conic composed of purely imaginary points on Π_∞ . Under the homography between Π_∞ and the image, the image of the absolute conic (IAC) is easily obtained from Lemma 3 as

$$\begin{aligned} \omega &= H_\infty^{-T} C_\infty H_\infty^{-1} = (KR)^{-T} I (KR)^{-1} \\ &= K^{-T} R^{-T} R^{-1} K^{-1} = (KK^T)^{-1} \end{aligned} \quad (4)$$

the dual image of the absolute conic (DIAC) is: $\omega^* = \omega^{-1} = KK^T$. It is clear that both the IAC and the DIAC depend only on the camera calibration matrix K .

Lemma 4. The image of the absolute conic ω is a symmetric matrix with five degrees of freedom (this is because the IAC is defined up to a scale). For a camera with rectangle pixels, it is easy to verify that $\omega_{12} = \omega_{21} = 0$, this can provide one linear constraint on the IAC. For a camera with square pixels, we have $\omega_{12} = \omega_{21} = 0$ and $\omega_{11} - \omega_{22} = 0$, this can provide two linear constraints on the IAC.

Lemma 5. Let v_1, v_2 be the vanishing points of two space lines, θ the included angle of the two lines, then

$$\cos \theta = \frac{v_1^T \omega v_2}{\sqrt{v_1^T \omega v_1} \sqrt{v_2^T \omega v_2}}$$

Lemma 6. If the two lines in Lemma 5 are orthogonal, then $v_1^T \omega v_2 = 0$, i.e. the vanishing points of the lines with orthogonal directions are conjugate with respect to the IAC.

Proofs of the above six lemmas can be found in [1,18]. From Lemma 6 we see that each pair of orthogonal vanishing points can provide one linear constraint on the IAC. If five such independent constraints can be obtained from the image, then the IAC can be computed linearly, and the intrinsic parameters of the camera can in turn be recovered straightforwardly from the IAC by Cholesky decomposition [1].

3. Camera calibration from a single view

3.1. Calibration from vanishing points and scene constraints

Usually, we can obtain three mutually orthogonal pairs of parallel lines from only one image of many man-made objects, such as architectures. Consequently, the three orthogonal vanishing points, say v_x, v_y, v_z , can be computed easily, as shown in Fig. 1. Therefore, three linear constraints on the IAC are obtained from only a single view of the scene. Most researchers use these constraints to calibrate the camera under the assumption of square pixels. Since under this assumption, two additional constraints can be obtained from Lemma 4, thus, the image of the absolute conic can be computed linearly. In this case, it was also shown that the camera's principal point is the orthocenter of the triangle with the three orthogonal vanishing points as vertices [1,15].

The square-pixel assumption, however, is much less tenable and does not hold for most off-the-shelf digital cameras, then the above algorithm may fail or give a poor solution. A question arises here: Can some additional constraints on the IAC be extracted from a single view?

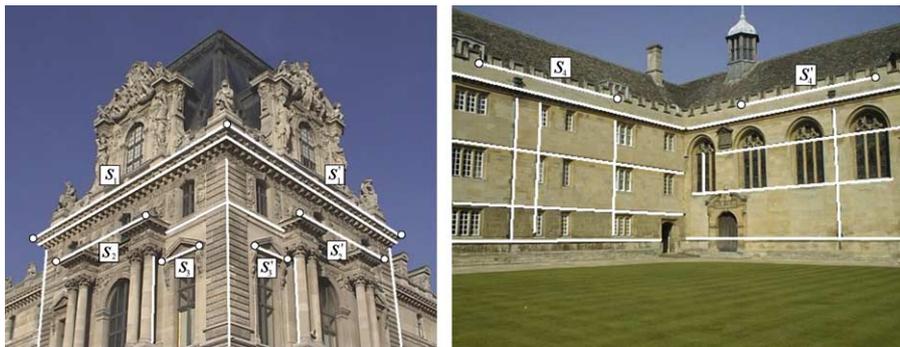


Fig. 1. Three orthogonal vanishing points can be obtained from a single image. There are some pairs of image segments whose corresponding space segments are of equal length, such as $s_1 \cong s'_1, s_2 \cong s'_2, s_3 \cong s'_3, s_4 \cong s'_4$.

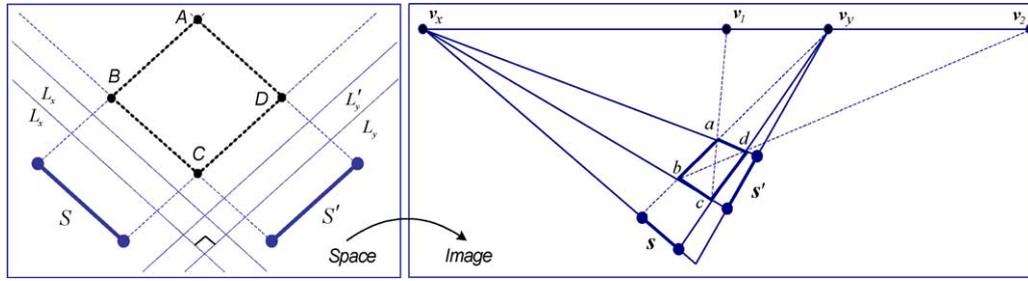


Fig. 2. v_x and v_y are the vanishing points of two mutually orthogonal pairs of parallel lines, respectively, two equal-length segments S and S' are parallel to the two pairs of lines. Then the formed quadrangle $abcd$ in the image must correspond to a square in space and the two intersection points v_1 and v_2 must be a pair of orthogonal vanishing points.

The answer is positive in most cases. This is because most of the man-made objects have some property of symmetry or contain some line segments with equal length or known length ratio. Let us see Fig. 1 again, it is easy to find that line segments $s_1 \cong s'_1, s_2 \cong s'_2, s_3 \cong s'_3, s_4 \cong s'_4$. We will show, in the following, how to use these properties in camera calibration.

Proposition 1. *Given two mutually orthogonal pairs of parallel lines L_x, L'_x and L_y, L'_y in the space (as shown in Fig. 2), two equal-length segments S and S' are parallel to L_x and L_y , respectively. Then, a new linear constraint on the image of absolute conic can be obtained.*

Proof. Suppose v_x is the vanishing point of L_x, L'_x, v_y is the vanishing point of L_y, L'_y . Then, the extension of image segments s and s' must pass through v_x and v_y , respectively. Connecting v_x with the endpoints of s', v_y with the endpoints of s forms a quadrangle $abcd$, which corresponds to $ABCD$ in the space, as shown in Fig. 2. It is easy to see that $AB \parallel CD, BC \parallel AD$ and $AB \perp BC$. Thus, quadrangle $ABCD$ is a square and its diagonal lines $AC \perp BD$. Suppose line ac intersects the vanishing line at v_1, bd intersects the vanishing line at v_2 , then v_1 and v_2 must be the vanishing points of two orthogonal directions. So we have $v_1^T \omega v_2 = 0$ from Lemma 6. \square

From v_1 and v_2 , together with the three orthogonal vanishing points v_x, v_y, v_z , four linearly independent constraints are obtained from a single view

$$\begin{cases} v_x^T \omega v_y = 0 \\ v_y^T \omega v_z = 0 \\ v_z^T \omega v_x = 0 \\ v_1^T \omega v_2 = 0 \end{cases} \quad (5)$$

If we can retrieve another set of orthogonal vanishing points on the vanishing line $v_y v_z$ or $v_x v_z$ from other sets of line segments, then five independent constraints can be obtained, and ω can be computed linearly. Unfortunately, this is rarely the case. Nevertheless, considering that most cameras can be assumed to have rectangle pixels (zero-skew), this is a quite natural and safe assumption for most imaging conditions, hence ω can still be solved linearly by the additional independent constraint $\omega_{12} = \omega_{21} = 0$

from Lemma 4. The constraint can also be written as $[1 \ 0 \ 0] \omega [0 \ 1 \ 0]^T = 0$, which means that the two axes of the image coordinates are orthogonal [15].

Remark 1. If there are more pairs of equal-length segments available in the scene, as shown in the left image of Fig. 1, then more squares can be obtained, and all the squares should be parallel in space. In this case, the vanishing points v_1 and v_2 can be computed by maximum likelihood estimation or other estimation methods so as to obtain a more faithful result.

Remark 2. If the line segments in Proposition 1 are with known length ratio rather than of equal length, then the angle between the diagonals AC and BD can be computed from the length ratio. In this case, v_1 and v_2 can also provide a constraint on the IAC from Lemma 5, but the constraint is non-linear. We will show in Section 3.2 that it can be converted into a linear one if the imaged circular points are recovered.

Remark 3. Degeneration will occur if one or more of the four points v_x, v_y, v_1, v_2 is located at infinity or near infinity in the image. Zisserman [15] gives a good discussion on the degeneration and ambiguities arising in camera calibration.

3.2. Calibration by virtue of circular points

In this section, the problem is further studied from the viewpoint of the circular points. We first show how to recover the images of the circular points from the two segments, then present three equivalent forms of constraints obtained from the imaged circular points and a vertical vanishing points. Unlike what is stated in Remark 2, all the computation here can be achieved linearly even the two segments are not equal in length.

Proposition 2. *The imaged circular points can be computed from the conditions given in Proposition 1. The same is true if the line segments in Proposition 1 are with known length ratio instead of equal length in space.*

Proof. When the two segments s and s' are with known length ratio in space, we can suppose the angle between the corresponding space lines of ac and bd is θ , the imaged circular points are m_i and m_j on the vanishing line, as shown

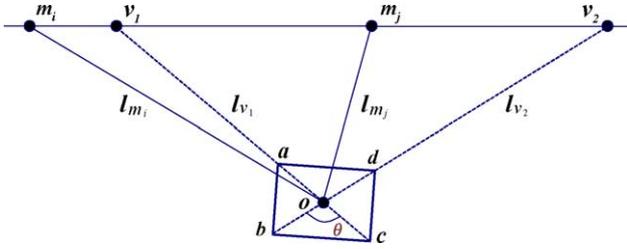


Fig. 3. The cross ratio of the four concurrent lines $l_{v1}, l_{v2}, l_{m1}, l_{m2}$, or four collinear points v_1, v_2, m_1, m_2 can be computed from the angle θ via Laguerre theorem.

in Fig. 3 (refer to Fig. 2). There are four concurrent lines $l_{v1}, l_{v2}, l_{m1}, l_{m2}$ through the intersection o of the lines ac and bd . From Laguerre theorem [19], we have

$$\theta = \frac{1}{2i} \ln(l_{v1}l_{v2}; l_{m1}l_{m2}) \quad (6)$$

where $(l_{v1}l_{v2}; l_{m1}l_{m2})$ is the cross ratio of the four concurrent lines. It is easy to verify that the cross ratio is equal to that of the four collinear points v_1, v_2, m_1, m_2 , i.e. $(l_{v1}l_{v2}; l_{m1}l_{m2}) = (v_1v_2; m_1m_2)$. Thus, Eq. (6) can be written as:

$$(v_1v_2; m_1m_2) = \exp(2\theta i) \quad (7)$$

When the two segments are of equal length, Eq. (7) becomes $(v_1v_2; m_1m_2) = -1$. On the other hand, since two orthogonal vanishing points are harmonic with the two imaged circular points m_i and m_j , we have:

$$(v_xv_y; m_im_j) = -1 \quad (8)$$

From (7) and (8), we can obtain two quadric equations which will provide two pairs of solutions of m_i and m_j , but only one solution is the true circular points (please refer to [16,17] for the proof). However, it is tedious to solve second order equations. We will introduce a linear method in the following.

From the above discussion, we know that the quadrangle $abcd$ corresponds to a rectangle (square) in the space, suppose the rectangle is $ABCD$ as shown in Fig. 4. Set the origin of the world coordinate system on the center of the rectangle with X and Y axes parallel to the two sides of

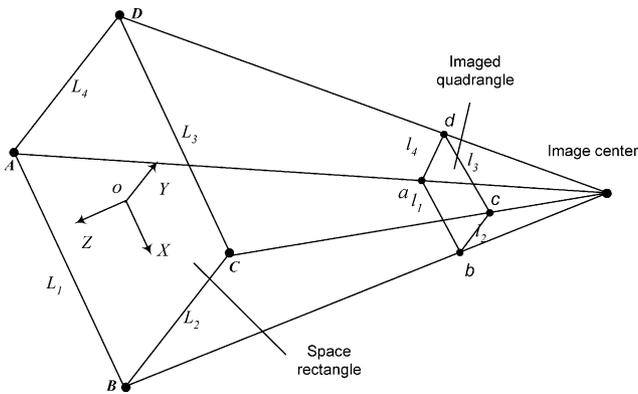


Fig. 4. The homography between the space rectangle and the image plane can be computed from four line correspondences $L_i \leftrightarrow l_i$ ($i=1, \dots, 4$).

the rectangle, respectively. Then the coordinates of the four side lines L_i ($i=1, \dots, 4$) can be computed easily as:

$$\begin{cases} L_1 = [0, 1, \sin \frac{\theta}{2}]^T, & L_2 = [1, 0, -\cos \frac{\theta}{2}]^T \\ L_4 = [1, 0, \cos \frac{\theta}{2}]^T, & L_3 = [0, 1, -\sin \frac{\theta}{2}]^T \end{cases} \quad (9)$$

Therefore, we have four line correspondences $L_i \leftrightarrow l_i$ ($i=1, \dots, 4$), the homography H between the space plane and the image can be linearly computed via SVD or through the following equation

$$H = [l_1, l_2, l_3]^{-T} \text{diag}(d_1, d_2, d_3) [L_1, L_2, L_3]^T \quad (10)$$

where

$$d_j = \frac{([L_1, L_2, L_3]^{-1} l_4)_j}{([l_1, l_2, l_3]^{-1} L_4)_j}, \quad j = 1, 2, 3,$$

$(a)_j$ denotes the j th element of vector a . The canonical form of the circular point in the space plane is $[1, \pm i, 0]^T$. Thus, the imaged circular points can be uniquely computed as:

$$\begin{cases} m_i \approx H[1, +i, 0]^T \\ m_j \approx H[1, -i, 0]^T \end{cases} \quad (11)$$

The images of the circular points are independent with the selection of the coordinate system. Thus, we may also select other world coordinate system. \square

Proposition 3. *Four independent linear constraints on the IAC can be obtained from the images of circular points m_i, m_j , and a vertical vanishing point v_z , which is the image of the direction perpendicular to parallel planes going through m_i and m_j .*

Proof. Since the imaged circular points are a pair of complex conjugate points lying on the image of absolute conic, they satisfy the following equations:

$$\begin{cases} m_i^T \omega m_i = 0 \\ m_j^T \omega m_j = 0 \end{cases} \quad (12)$$

On the other hand, $m_i(m_j)$ and v_z can be considered as the vanishing points of lines with orthogonal directions. So from Lemma 6 we have:

$$\begin{cases} m_i^T \omega v_z = 0 \\ m_j^T \omega v_z = 0 \end{cases} \quad (13)$$

It is clear that (12) and (13) altogether provide four linearly independent constraints on ω . \square

Since m_i and m_j are a pair of complex conjugate 3-vectors, let vector v_r and v_i be the corresponding real and imaginary parts, i.e. $m_i = v_r + iv_i, m_j = v_r - iv_i$, then the constraints in (12) and (13) can be written as the following

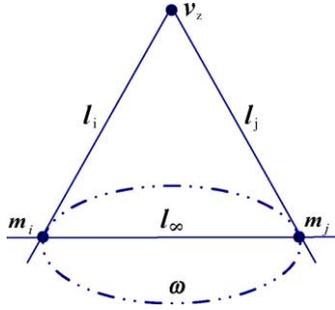


Fig. 5. The pole–polar relationship with respect to the IAC between the vanishing line l_∞ and the vertical vanishing point v_z .

equivalent form through a simple computation

$$\begin{cases} v_r^T \omega v_z = 0 \\ v_i^T \omega v_z = 0 \\ v_r^T \omega v_r - v_i^T \omega v_i = 0 \\ v_r^T \omega v_i + v_i^T \omega v_r = 0 \end{cases} \quad (14)$$

Another set of equivalent constraints can also be obtained as follows. Since there is a pole–polar relationship with respect to the IAC between the vanishing line $l_\infty = m_i \times m_j$ and the vertical vanishing point v_z (see Fig. 5), lines $l_i = m_i \times v_z$ and $l_j = m_j \times v_z$ are tangent to the IAC at the circular points m_i and m_j . Thus, we have

$$\begin{cases} m_i \times m_j = \lambda_1 \omega v_z \\ m_i \times v_z = \lambda_2 \omega m_i \\ m_j \times v_z = \lambda_3 \omega m_j \end{cases} \quad (15)$$

where $\lambda_1, \lambda_2, \lambda_3$ are unknown non-zero scales. Each equation in (15) can provide two linear constraints on the IAC, but only four of them are independent, which are equivalent to those given in (14). Actually, the last two equations in (15) provide the same constraints, since m_i and m_j are complex conjugate.

We have obtained three sets of linear constraints which are equivalent in essence. In practice, we can use one of them so as to simplify the computation. From the four independent constraints, a one-parameter family of solutions for the IAC can be obtained, while this ambiguity can also be solved by the assumption of rectangle pixels as in Section 3.1.

4. Recovery of extrinsic parameters and camera projection matrix

Proposition 4. A rotation matrix and projection matrix with respect to the world coordinate system of Fig. 4 can be recovered given the camera's calibration matrix.

Proof. We have computed the homography in Section 3.2. Then from Eq. (1) we have

$$sH = K(r_1, r_2, t), \quad \text{s.t. } \|r_1\| = \|r_2\| = 1 \quad (16)$$

where r_i is the i th column of the orthonormal rotation matrix R , t is the translation vector. Let $K^{-1}H = [a_1, a_2, a_3]$. It is easy to see that $s = \pm 1/\|a_1\| (= \pm 1/\|a_2\|)$, $r_1 = sa_1$, $r_2 = sa_2$, $r_3 = r_1 \times r_2$, $t = sa_3$. Thus, we have two pairs of solutions of R and t as:

$$R_1 = \begin{bmatrix} a_1 & a_2 & a_1 \times a_2 \\ \frac{1}{\|a_1\|} & \frac{1}{\|a_2\|} & \frac{1}{\|a_1 \times a_2\|} \end{bmatrix}, \quad t_1 = \frac{a_3}{\|a_1\|} \quad (17)$$

$$R_2 = \begin{bmatrix} -a_1 & -a_2 & a_1 \times a_2 \\ \frac{1}{\|a_1\|} & \frac{1}{\|a_2\|} & \frac{1}{\|a_1 \times a_2\|} \end{bmatrix}, \quad t_2 = -\frac{a_3}{\|a_1\|} \quad (18)$$

It is obviously that only one pair of the above solutions can make the reconstructed objects lie in the front of the camera, this is the correct solution. The camera projection matrix can then be computed straightforwardly from Eq. (1) as $P = K[R, t]$. \square

5. Measurement and 3D reconstruction from a single view

As three orthogonal vanishing points have been obtained, the scene in the world may be taken in general as in the shape of a cuboid. Let us assume the world coordinate system is on the cuboid (see Fig. 6 and note that other selections are optional). Then the camera projection matrix can be retrieved with respect to the world system. In this section, we will present a method of 3D reconstruction from a single view based on scene measurement.

Most man-made objects, especially architectures, are usually composed of many pieces of planar surfaces. If the world coordinates of each surface can be obtained, then all the 3D information on the surface can be recovered accordingly. Thus, the whole object is assembled by merging the planar patches into a 3D structure [14,18].

Suppose the coordinates of a space plane is Π_i , i.e. $\Pi_i^T x = 0$, with $x = [x, y, z, 1]^T$, then for an image point m_j on the plane, its corresponding space point x_j can be easily computed by the intersection of the back-projected line

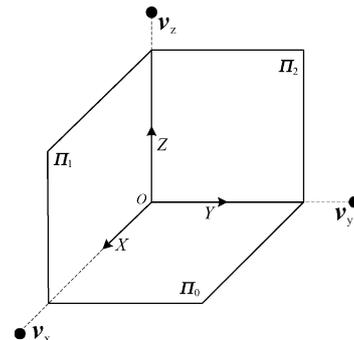


Fig. 6. The world coordinate system and three orthogonal vanishing points.

and the plane as:

$$\begin{cases} sm_j = Px_j \\ \Pi_i^T x_j = 0 \end{cases} \quad (19)$$

Similarly, for an image line l , its back-projection is a plane $\Pi = P^T l$; for a conic C in the image, its back-projection is a view cone $Q = P^T C P$. Their corresponding space coordinates can also be computed linearly by the intersection of the back-projection and the space plane.

In Fig. 6, the coordinates of plane Π_0 , Π_1 and Π_2 are obvious, while the coordinates of most other planar surfaces can be recovered from the scene constraints with respect to the three base planes [14]. Let us take the plane Π_0 as an example.

Proposition 5. *The coordinates of plane Π_i parallel to Π_0 can be retrieved if a pair of corresponding points on Π_0 and Π_i in the direction of v_z can be obtained from the image.*

Proof. To recover the coordinates of Π_i is equivalent to retrieving the distance z_0 between Π_0 and Π_i . Suppose $x = [x_0, y_0, 0, 1]^T$ and $x' = [x_0, y_0, z_0, 1]^T$ are the pair of corresponding points on the two planes, with three unknowns x_0, y_0, z_0 , as shown in Fig. 7. Their corresponding images are $m_x = [u_x, v_x, 1]^T$ and $m'_x = [u'_x, v'_x, 1]^T$, respectively. Then from

$$\begin{cases} s_1 m_x = Px \\ s_2 m'_x = Px' \end{cases} \quad (20)$$

it is easy to obtain the coordinates of plane $\Pi_i = [0, 0, 1, -z_0]^T$. We can also use this proposition to compute the height of an object on the plane Π_0 . □

Proposition 6. *Suppose an arbitrary plane Π_a intersect Π_0 at line L (see Fig. 8), then the coordinates of plane Π_a can be determined from the images of a pair of parallel lines on the plane.*

Proof. Since L lies on the plane $\Pi_0 = [0, 0, 1, 0]^T$, its coordinates can be easily computed. Suppose $L = [a, b, 0, d]^T$, Π_v is the plane passing through L and perpendicular to Π_0 , then Π_v must have the same coordinates as L . All the planes passing through L form a pencil, and the pencil can be expressed as $\Pi_a = \Pi_v + \lambda \Pi_0 = [a, b, \lambda, d]^T$, with λ the only unknown parameters here.

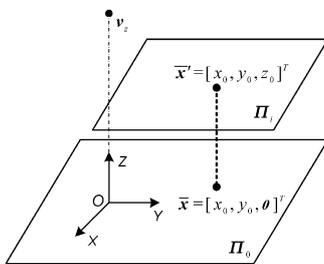


Fig. 7. The vertical vanishing point v_z and a pair of corresponding points between two parallel planes.

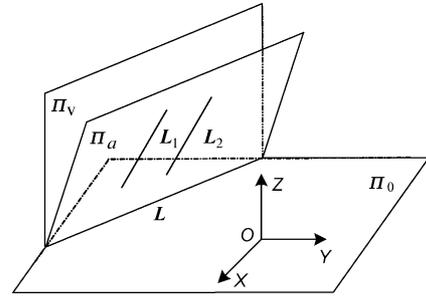


Fig. 8. The relationship between the reference plane, vertical plane and an arbitrary plane. There is a pencil of planes passing through line L .

Denote the parallel lines in space and their corresponding images as L_1, L_2 and l_1, l_2 , respectively, the back-projection of l_1, l_2 form two space planes $\Pi_{b1} = P^T l_1$ and $\Pi_{b2} = P^T l_2$. Denote the normal vector of plane $\Pi_a, \Pi_{b1}, \Pi_{b2}$ as n_a, n_{b1}, n_{b2} , let the direction vector of L_1, L_2 be n_{L1}, n_{L2} , then $n_{L1} = n_a \times n_{b1}$, $n_{L2} = n_a \times n_{b2}$. From $L_1 // L_2$, λ can be easily computed via the least squares. Thus, the coordinates of Π_a is recovered. □

Remark 4. If some other prior information in the arbitrary plane, such as two orthogonal lines, the coordinates of a point, etc. can be retrieved from the image, the scalar λ , as well as the coordinates of the arbitrary plane, can also be computed in a similar way.

Remark 5. Other geometrical entities, such as the distance between two lines, distance from a point to a plane, angle formed by two lines or two planes, angle formed by a line and a plane, etc. can also be recovered by combining the scene constraints [18].

6. Experiments with simulated data

During the simulations, we generate a cube in the space, whose size and position in the world coordinate system are shown in Fig. 9. There are three parallel lines corresponding to each direction of the three world axes, each line is

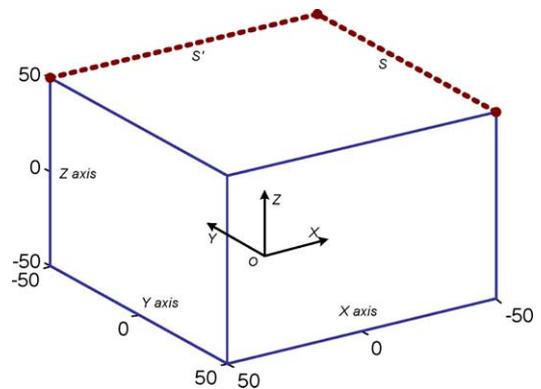


Fig. 9. The simulated cube and the world coordinate system, there are three pair of parallel lines corresponding to the directions of three axes of the world system.

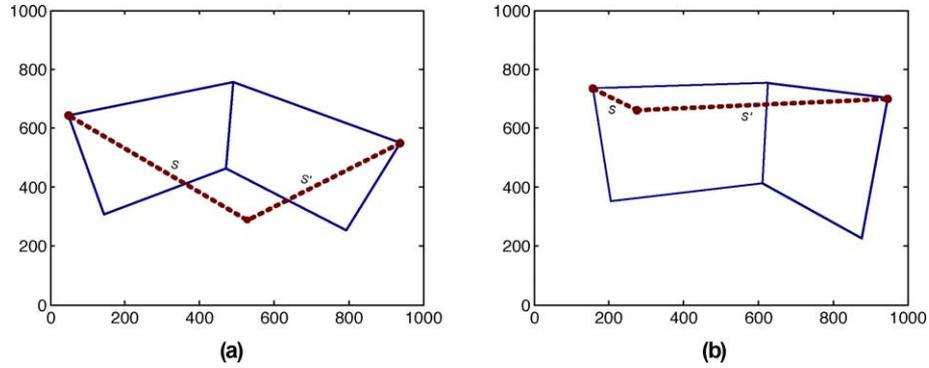


Fig. 10. The image of the simulated cube in (a) Case 1 and (b) Case 2. Note that the two segments s and s' correspond to segments with equal length in space, but their lengths in images are quite different due to different imaging conditions.

Table 1
The comparative calibration results of our method (RA) and Wilczkowiak's method (PA) in Case 1

Noise level (δ)			0.4	0.8	1.2	1.6	2.0	2.4	2.8	3.2	3.6
f_u	Mean	RA	0.019	0.034	0.065	0.160	0.467	0.788	1.028	1.579	1.948
		PA	0.028	0.049	0.121	0.418	0.928	1.407	1.964	2.447	2.809
	STD	RA	0.112	0.199	0.413	0.722	0.928	1.342	1.822	2.340	2.908
		PA	0.149	0.237	0.521	0.947	1.143	1.408	1.842	2.292	2.819
f_v	Mean	RA	0.018	0.028	0.053	0.165	0.453	0.714	0.978	1.465	1.727
		PA	0.021	0.044	0.117	0.366	0.847	1.097	1.873	2.234	2.720
	STD	RA	0.078	0.153	0.334	0.657	0.884	1.157	1.699	2.197	2.660
		PA	0.093	0.204	0.431	0.797	1.054	1.306	1.735	2.441	3.046
u_0	Mean	RA	0.009	0.025	0.045	0.141	0.377	0.587	0.918	1.357	1.515
		PA	0.013	0.039	0.103	0.325	0.795	1.050	1.653	2.151	2.585
	STD	RA	0.061	0.115	0.286	0.570	0.845	1.070	1.410	1.941	2.202
		PA	0.088	0.184	0.362	0.658	0.905	1.134	1.501	2.107	2.640
v_0	Mean	RA	0.028	0.043	0.086	0.214	0.543	0.931	1.326	1.731	2.174
		PA	0.041	0.081	0.204	0.346	0.962	1.439	2.014	2.496	2.961
	STD	RA	0.132	0.238	0.481	0.863	0.992	1.402	1.845	2.584	2.970
		PA	0.159	0.251	0.569	0.926	1.094	1.448	1.969	2.426	2.888

In each noise level, 500 independent tests are taken. We can see from the mean and STD of the relative errors (%) of four intrinsic parameters that our method performs better than that of Wilczkowiak's in this case.

Table 2
The comparative calibration results of our method (RA) and Wilczkowiak's method (PA) in Case 2, Wilczkowiak's method performs better than that of ours in this case

Noise level (δ)			0.4	0.8	1.2	1.6	2.0	2.4	2.8	3.2	3.6
f_u	Mean	RA	0.056	0.109	0.346	0.917	2.026	2.562	3.318	4.209	5.129
		PA	0.036	0.069	0.207	0.576	1.270	1.766	2.404	3.168	3.795
	STD	RA	0.328	0.630	1.022	1.771	2.189	2.576	3.614	4.614	5.617
		PA	0.208	0.436	0.662	1.266	1.586	1.901	2.588	3.371	4.381
f_v	Mean	RA	0.050	0.086	0.299	0.805	1.898	2.306	2.973	3.867	4.976
		PA	0.029	0.064	0.175	0.557	1.136	1.573	2.222	2.955	3.628
	STD	RA	0.274	0.461	0.896	1.714	1.998	2.547	3.336	4.310	5.443
		PA	0.152	0.369	0.606	1.183	1.437	1.732	2.345	3.044	4.171
u_0	Mean	RA	0.046	0.118	0.261	0.702	1.531	2.029	2.099	3.643	4.648
		PA	0.021	0.068	0.153	0.462	0.897	1.525	2.118	2.752	3.339
	STD	RA	0.205	0.423	0.752	1.586	1.930	2.000	3.164	3.884	4.847
		PA	0.121	0.285	0.488	1.074	1.391	1.728	2.137	2.833	3.878
v_0	Mean	RA	0.043	0.133	0.525	1.152	2.187	2.870	3.608	4.509	5.627
		PA	0.041	0.084	0.272	0.739	1.495	1.949	2.512	3.488	4.183
	STD	RA	0.349	0.675	0.880	2.106	2.254	2.932	3.783	4.765	5.804
		PA	0.230	0.514	0.667	1.439	1.773	2.118	2.803	3.653	4.501

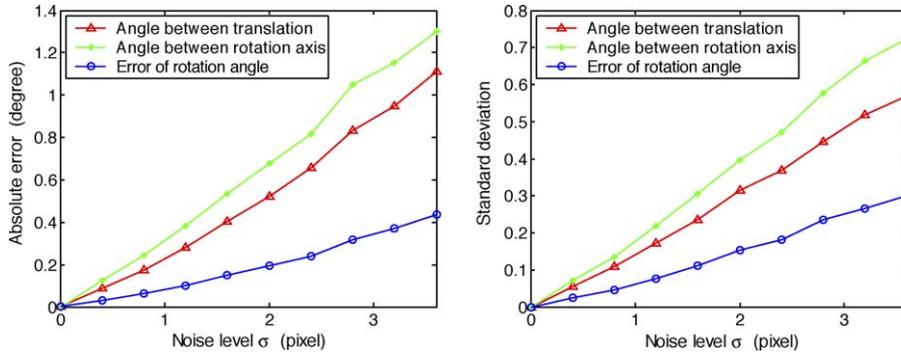


Fig. 11. The absolute errors and standard deviations of translation vector, rotation axis and rotation angle.

composed of 100 evenly distributed points. Gaussian image noise (with mean zero) is added on each imaged point, and the corresponding image lines are fitted from these points using the least squares fitting. The vanishing points are computed as the intersection of each set of parallel lines. We also select the two space line segments (in dashed lines in Fig. 9) S and S' of equal length to do calibration using our proposed method.

6.1. Comparative tests for calibration of intrinsic parameters

Case 1. The setup of the camera is: $f_u = 1200, f_v = 1000, s = 0, u_0 = 510, v_0 = 490$, image size is 1000×1000 pixels, rotation axis $r_1 = [0.6988, 0.7070, -0.1088]^T$, rotation angle $\alpha_1 = -60.805^\circ$ and translation vector $t_1 = [-10, -20, 210]^T$. The image is shown in Fig. 10(a). In this case, the images of the three orthogonal vanishing points are $v_x = [2041, 1091]^T, v_y = [218, -655]^T, v_z = [-1084, 1645]^T$, respectively, and the estimated vanishing points may vary a little bit under different noise level. We use the method proposed in Section 3.2. Table 1 gives the means and standard deviations (STDs) of the relative errors of four estimated intrinsic parameters, respectively. We also give a comparative calibration result via Wilczkowiak’s method [11] under the same condition. In

Table 1, RA stands for our method, PA for Wilczkowiak’s. The mean and STD in each noise level (the STD of the Gaussian noise, unit: pixel) are given by 500 independent tests in order to provide more statistically meaningful results.

Case 2. The intrinsic parameters of the camera are the same as in Case 1, while the rotation axis $r_2 = [-0.6576, -0.7419, 0.1308]^T$, rotation angle $\alpha_2 = 30.02^\circ$ and translation vector $t_2 = [0, 0, 220]^T$. The corresponding image is shown in Fig. 10(b). Note that the lengths of the two equal segments are quite different in this condition. The three computed vanishing points under noise level $\delta = 0$ are $v_x = [3593, 854]^T, v_y = [510, -2258]^T, v_z = [-19, 854]^T$, respectively. The comparative calibration results with Wilczkowiak’s method (PA) are shown in Table 2.

From the above tests, as well as some other tests, we find that in some cases, as in the test of Case 1, our proposed method has lower relative errors and STDs in camera calibration than Wilczkowiak’s method via parallelepiped. While in other cases, as in the test of Case 2, Wilczkowiak’s method performs better than ours, but the accuracy of our proposed method is still acceptable. One reason is that our proposed method is based on vanishing points, it may near to degenerate when the imaged vanishing points locate very far away from the images, or even at infinity (in Case 2, the positions of the estimated vanishing points are in general far away than those in Case 1). This degeneracy may happen in all vanishing points based methods. The other reason is that in Case 2, the lengths of the two line segments (which are of equal length in space and the constraint is used for

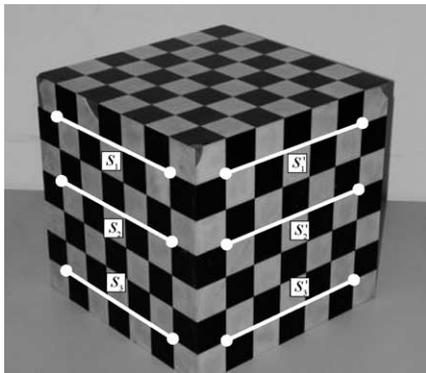


Fig. 12. One image of a calibration block with three pairs of equal-length segments used for calibration.

Table 3
The first test. A comparative calibration results by the proposed method (RA), the square-pixel assumption (SA) and the DLT method

	f_u	f_v	s	u_0	v_0	Mean	STD
RA	2198.1	2236.4	0	563.8	346.9	7.583	0.038
SA	2200.0	2200.0	0	579.7	274.6	7.699	0.049
DLT	2254.9	2272.6	-0.01	556.8	371.5	7.160	0.029

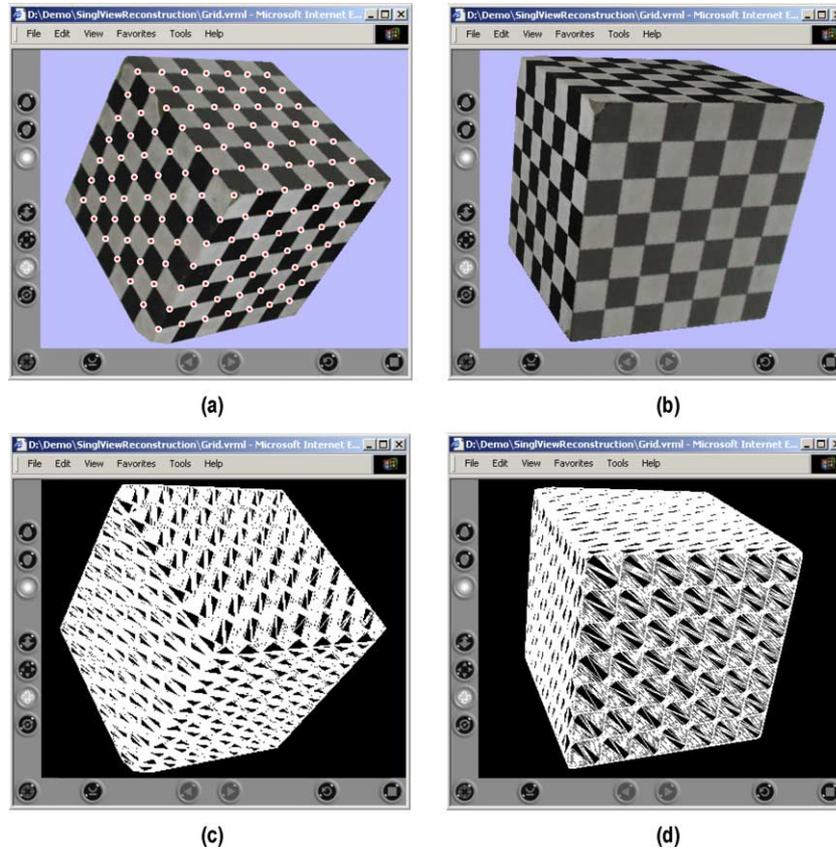


Fig. 13. (a) and (b) The reconstruction results of the calibration block under different viewpoints with texture mapping. (c) and (d) The corresponding reconstruction results in triangulated wire frames under different viewpoints.

calibration) are quite different in the image. This may cause relative more errors in the endpoints detection of the shorter segment. Therefore, we should avoid the degeneracy cases during image taking so as to increase the computational accuracy. In Case 2, if we select other equal-length space segments which are of small length difference in the image rather than s and s' , tests shows



Fig. 14. One image of a church in Valbonne with three pairs of equal-length segments used for calibration.

that the two methods are with comparative accuracy. This calibration results are omitted here due to space limitation.

One may have noted that Wilczkowiak’s method also performs much worse in Case 2 than in Case 1, though the method does not involve in the degeneracy of vanishing points. However, the method depends on the determination of three-dimension parameters and three angles of the parallelepiped. The estimation of these parameters may suffer from a loss of accuracy in Case 2, since the simulated cube is greatly deformed in the image (see Fig. 10(b)).

6.2. Tests for calibration of extrinsic parameters

Using the estimated intrinsic parameters in Case 1, the proposed method in Section 4 is used to compute

Table 4
The second and third tests. A comparative calibration results by the proposed method (RA) and the square-pixel assumption (SA) method

		f_u	f_v	s	u_0	v_0
Test 2	RA	720.86	756.22	0	255.48	379.04
	SA	724.31	724.31	0	260.72	386.78
Test 3	RA	1070.6	1118.2	0	523.8	378.3
	SA	1094.4	1094.4	0	565.1	373.7

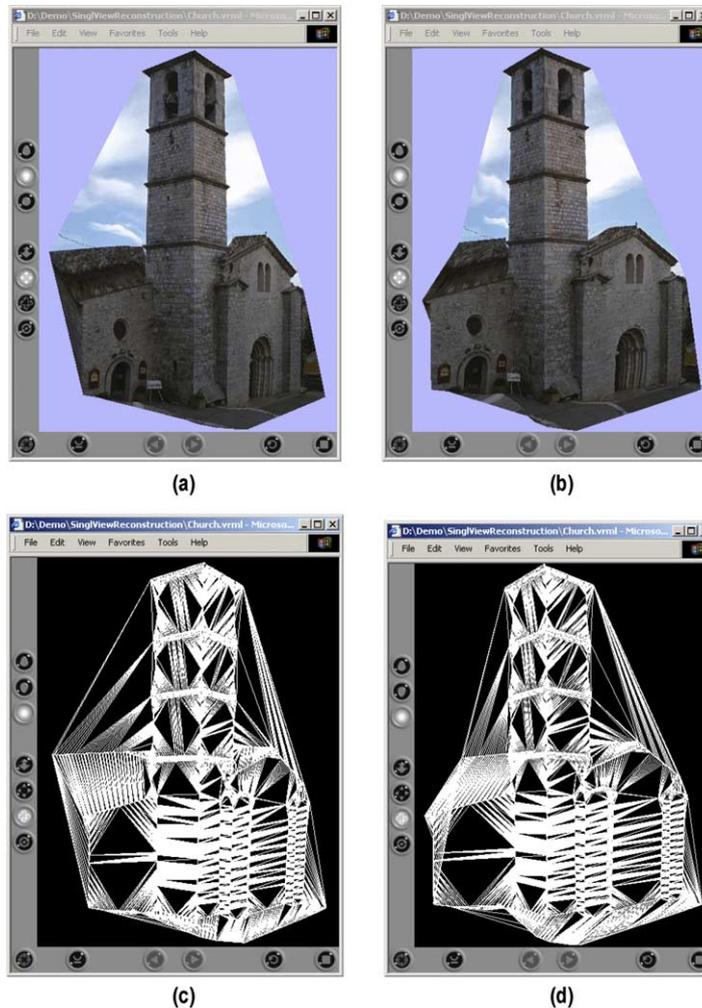


Fig. 15. (a) and (b) The reconstruction results of the church under different viewpoints with texture mapping. (c) and (d) The corresponding reconstruction results in triangulated wire frames under different viewpoints.

the extrinsic parameters of the camera. In order to facilitate the comparison, we decompose the estimated rotation matrix into a rotation axis and a rotation angle. Since the ground truth is known, we compare the angular error between the estimated and real rotation axis, the error of the rotation angle and the angular error between the estimated translation vector and the real one. The absolute errors and STDs of these entities are plotted in Fig. 11 (the values in each noise level are computed from 500 independent tests). From the results we can see that our proposed method gives very small absolute errors and STDs even under higher noise level when the skew of the camera is zero. If we use the calibration results in Case 2, the absolute errors and STDs in each noise level are a little bit higher than those in Case 1. The figures are omitted here due to space limitation.

7. Experiments with real images

The experiments are performed on three test images. The first test image is a calibration block, as shown in Fig. 12.

The image is taken by a Nikon Coolpix 990 digital camera with resolution of 1024×768 . During the test, we use Canny edge detector to detect the edge points and use Hough transform to fit the detected points into straight lines. Then, the vanishing points are computed from the parallel



Fig. 16. A image of the Wadham College of Oxford with one pairs of equal-length segments used for calibration.

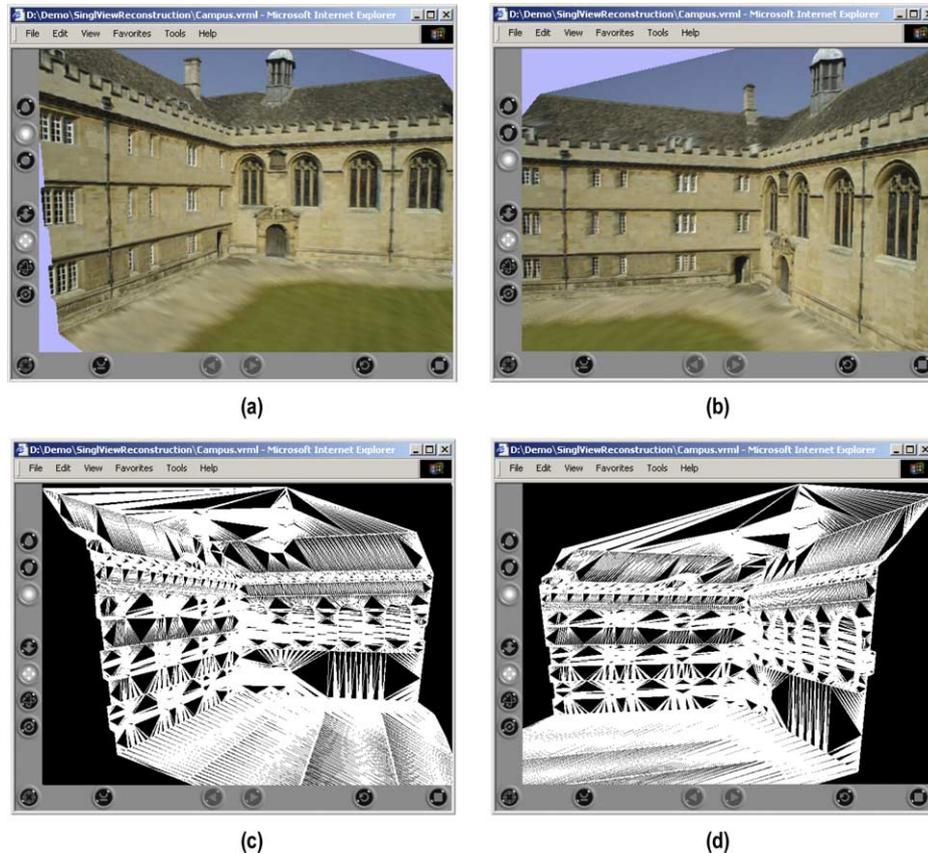


Fig. 17. (a) and (b) The reconstruction results of the Wadham College under different viewpoints with texture mapping. (c) and (d) The corresponding reconstruction results in triangulated wire frames under different viewpoints.

lines using a maximum likelihood estimator [1,7]. The images of circular points are computed from three selected pairs of equal-length segments (see Fig. 12) using the least squares fitting.

The calibration results are shown in Table 3. We also give a comparative test on using only three orthogonal vanishing points under the assumption of square pixels. Since the size of each square patches is known precisely for a calibration block, the camera can be calibrated via direct linear transform (DLT) with five intrinsic parameters [1,2]. The result is shown in Table 3, where, RA stands for our proposed method; SA the method under square-pixel assumption.

Fig. 13(a)–(d) gives the reconstruction results of Fig. 12 by the proposed method in Section 5. We also mark the reconstructed corner points of each square patches of the calibration block in Fig. 13(a). The distance between each pair of adjacent points should be equal theoretically. The mean value and STD of all the distances estimated by each method are also shown in Table 3. We can see from the results that the proposed method is better than those of only using three vanishing points. The results of our method are close to those given by DLT, which can only be applied to the cases where more space points are precisely located.

The second test image is a Church in Valbonne, as shown in Fig. 14, which is downloaded from the Visual Geometry Group of the University of Oxford. The image

resolution is 512×768 . We use the same method as in the first test to detect the three orthogonal vanishing points, and compute the image of circular points via three pairs of symmetric line segments in the chimney. The comparative calibration results are listed in Table 4. The reconstruction results under different viewpoint with texture mapping and in triangulated wire frames are shown in Fig. 15(a)–(d).

The third test image is the Wadham College of Oxford, as shown in Fig. 16, which is also downloaded from the Visual Geometry Group, with the resolution of 1024×768 . The comparative calibration results are shown in Table 4. The reconstruction results are given in Fig. 17(a)–(d). Note that the depth of each window and door are recovered by the proposed method.

We can see from all the reconstructions that they are largely consistent with the real cases, and seem very realistic.

8. Conclusions

In this paper, we mainly focus on the problem of camera calibration and 3D reconstruction from a single view of a structured scene. We propose and prove that two line segments with equal length or known length ratio in

the scene can provide an additional independent constraint to the image of the absolute conic. The constraint is expressed both in terms of a pair of orthogonal vanishing points and the image of circular points. This is a development to the popular calibration method based on three orthogonal vanishing points. We also present a simple method for the recovery of camera extrinsic parameters and projection matrix with respect to a given world system. Furthermore, we extend the single view metrology [14] to the Euclidean space to estimate the position and pose of a space planar surface from the recovered projection matrix and scene constraints. Thus, the scene structure can be reconstructed by combining planar patches. The recovery is very effective as fine details like the depths of windows in a building can also be found, since scene constraints are utilized during the reconstruction. Extensive experiments on simulated data and real images validate our proposed approach and show that the method is better than the method using only three vanishing points. A comparative experimental study with other method [11] is also performed. In many scenes, such as those containing buildings and other man-made objects, the additional constraint can be found. Thus, the proposed approach will have wide applications in 3D modeling. It is a simple and convenient method using a single image and the difficult matching problem has been avoided at the expense of minimal human interaction.

It is clear that the proposed methods are based on some known specific geometrical information about the scene, and the precision of the approach depends greatly on the image preprocessing, such as edge detection, line fitting and vanishing point detection. Hence, it is crucial to select a robust algorithm for vanishing points computation so as to improve the accuracy of calibration and reconstruction [7, 20,21]. It should be noted that lens distortion has not been considered in our methods, since the camera used in our experiment has negligible lens distortion. However, the image should be rectified firstly if the lens distortion does play a significant role in the accuracy of measurement.

Acknowledgements

The authors would like to thank the reviewers for their valuable comments and suggestions. Thanks to the Visual Geometry Group of the University of Oxford for providing the test images. The work is supported by the Hong Kong RGC Grant CUHK 4378/02E, the National Natural Science Foundation of China under grant no. 60175009 and the National High Technology Development program of China under grant no. 2002AA135110.

References

- [1] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, MA, 2000.
- [2] O.D. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*, MIT Press, Cambridge, MA, 1993.
- [3] Y. Horry, K. Anjyo, K. Arai, Tour into the picture, in: *Proceedings of SIGGRAPH*, 1997, pp. 225–232.
- [4] L. Zhang, G. Dugas-Phocion, J.S. Samson, S.M. Seitz, Single view modeling of free-form scenes, in: *Proceedings of Computer Vision and Pattern Recognition*, Kauai, Hawaii, December 2001, pp. 990–997.
- [5] P. Debevec, C.J. Taylor, J. Malik, Modeling and rendering architecture from photographs: a hybrid geometry- and imaged-based approach, in: *Proceedings of SIGGRAPH*, 1996, pp. 11–21.
- [6] B. Caprile, V. Torre, Using vanishing points for camera calibration, *International Journal of Computer Vision* 4 (2) (1990) 127–140.
- [7] D. Liebowitz, A. Criminisi, A. Zisserman, Creating architectural models from images, in: *Proceedings of Eurographics*, Milan, Italy, September 1999, pp. 39–50.
- [8] F.A. van den Heuvel, 3D Reconstruction from a single image using geometric constraints, *ISPRS Journal of Photogrammetry and Remote Sensing* 53 (6) (1998) 354–368.
- [9] D. Liebowitz, A. Zisserman, Combining scene and auto-calibration constraints, in: *Proceedings of International Conference on Computer Vision*, Kerkira, Greece, September 1999, pp. 285–292.
- [10] R. Cipolla, E. Boyer, 3D model acquisition from uncalibrated images, in: *Proceedings of IAPR Workshop on Machine Vision Applications*, Chiba, Japan, November 1998, pp. 559–568.
- [11] M. Wilczkowiak, E. Boyer, P. Sturm, Camera calibration and 3D reconstruction from single images using parallelepipeds, in: *Proceedings of International Conference on Computer Vision*, Vancouver, Canada, July 2001, vol. I, pp. 142–148.
- [12] M. Wilczkowiak, E. Boyer, P. Sturm, 3D modeling using geometric constraints: a parallelepiped based approach, in: *Proceedings of European Conference on Computer Vision*, LNCS2353, Copenhagen, Denmark, May, 2002, vol. IV, pp. 221–237.
- [13] C.S. Chen, C.K. Yu, Y.P. Hung, New calibration-free approach for augmented reality based on parameterized cuboid structure, in: *Proceedings of International Conference on Computer Vision*, Kerkira, Greece, September 1999, pp. 30–37.
- [14] A. Criminisi, I. Reid, A. Zisserman, Single view metrology, *International Journal of Computer Vision* 40 (2) (2000) 123–148.
- [15] A. Zisserman, D. Liebowitz, M. Armstrong, Resolving ambiguities in auto-calibration, *Philosophical Transactions of the Royal Society of London, Series A* 356 (1740) (1998) 1193–1211.
- [16] G.H. Wang, Z.Y. Hu, F.C. Wu, Novel approach to circular points based camera calibration, *Proceedings of SPIE*, vol. 4875, 2002, pp. 830–837.
- [17] F.C. Wu, G.H. Wang, Z.Y. Hu, A linear approach for determining intrinsic parameters and pose of cameras from rectangles, *Chinese Journal of Software (In Chinese)* 14 (3) (2003) 703–712.
- [18] G.H. Wang, Z.Y. Hu, F.C. Wu, Single view based measurement on space planes, *Journal of Computer Science and Technology* 19 (3) (2004).
- [19] J.G. Semple, G.T. Kneebone, *Algebraic Projective Geometry*, Oxford University Press, Oxford, 1952.
- [20] G. McLean, D. Kotturi, Vanishing point detection by line clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7 (11) (1995) 1090–1095.
- [21] F.A. van den Heuvel, Vanishing point detection for architectural photogrammetry, in: *International Archives of Photogrammetry and Remote Sensing*, Hakodate, Japan, 1998, pp. 652–659.
- [22] S.K. Nayar, Y. Nakagawa, Shape from focus, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (8) (1994) 824–831.
- [23] B.J. Super, A.C. Bovik, Shape from texture using local spectral moments, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17 (4) (1995) 333–343.