Available online at www.sciencedirect.com
ScienceDirect
Computer Vision and Image Understanding 105 (2007) 60–72
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Camera calibration and light source orientation from solar shadows

Xiaochun Cao *, Hassan Foroosh

Computational Imaging Laboratory, University of Central Florida, Orlando, FL 32816-2362, USA

Received 2 March 2006; accepted 10 August 2006
Available online 26 September 2006

Abstract

In this paper, we describe a method for recovering camera parameters from perspective views of daylight shadows in a scene, given only minimal geometric information determined from the images. This minimal information consists of two 3D stationary points and their cast shadows on the ground plane. We show that this information captured in two views is sufficient to determine the focal length, the aspect ratio, and the principal point of a pinhole camera with fixed intrinsic parameters. In addition, we are also able to compute the orientation of the light source. Our method is based on exploiting novel inter-image constraints on the image of the absolute conic and the physical properties of solar shadows. Compared to the traditional methods that require images of some precisely machined calibration patterns, our method uses cast shadows by the sun, which are common in natural environments, and requires no measurements of any distance or angle in the 3D world. To demonstrate the accuracy of the proposed algorithm and its utility, we present the results on both synthetic and real images, and apply the method to an image-based rendering problem.

Keywords: Camera calibration; Solar shadow; Light source orientation estimation; Planar homology

1. Introduction

Camera calibration is an essential step in computer vision for 3D Euclidean reconstruction and motion estimation. It also plays a crucial role in applications such as augmented reality, and production of special effects. Techniques from augmented reality and inverse rendering now permit the rendering of synthetic models within images or video sequences. However, in order for these techniques to be successful, cameras need to be calibrated to ensure that real and computer generated objects blend in a geometrically correct fashion in their mixed environment. On the other hand, to create realistic and seamless images, virtual objects have to be shaded consistently under the real illumination condition of the target scene, which in particular requires the knowledge of the light source direction. There is, therefore, an increasing need for calibration and light source estimation directly by using minimal information from the images of the target scene.

Traditional camera calibration methods use a calibration object with a fixed 3D geometry [10,15,44]. Recently, more flexible plane-based approaches [11,12,23,26,40,50] have been proposed that use the orthonormal properties of the rotation matrix, which represents the relative rotation between the 3D world and the camera coordinate system. Zhang [50] has shown that it is possible to calibrate a camera using a planar pattern viewed at a few different orientations. Sturm and Maybank [40] extended this technique, and obtained the solution for the case when the camera intrinsic parameters can vary. Gurdjos et al. [11] formulated the plane-based calibration problem in a more intuitive geometric framework according to Poncelet's theorem, which makes it possible to give a Euclidean interpretation to the plane-based calibration techniques. Other special objects (see Fig. 1), e.g. balls or spheres [1,41,49], surfaces of revolution [8,46], circles [7,18,28,47], non-textured Lambertian surfaces [19], fixed stars in the night sky [20] and 1D objects [51], have also been used in the past

* Corresponding author.
E-mail addresses: xccao@cs.ucf.edu (X. Cao), foroosh@cs.ucf.edu (H. Foroosh).

1077-3142/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2006.08.003

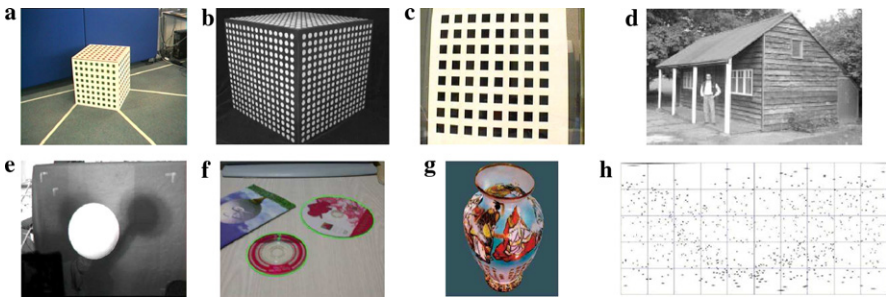

Fig. 1. A summary list of previous camera calibration objects. (a) 3D square grid [44]. (b) 3D circular grid [15]. (c) 2D square grid [40,50]. (d) Architectural buildings [23]. (e) Balls or spheres [1,41,49]. (f) Coplanar circles [7,18,28,47]. (g) Surfaces of revolution [8,46]. (h) Fixed stars in the night sky [20]. Note that this figure is by no means complete and only representative work is listed.

as alternative calibration objects. Although the recent auto-calibration techniques [14,27,33,38] avoid the onerous task of using special calibration objects, they mostly require more than three views and also assume a simplified camera model.

The calibration technique introduced in this paper, namely, calibration from shadows, is closely related to the existing methods [6,8,21,23,25,40,46,50] that explore the constraints provided by vanishing points along orthogonal directions, or the pole–polar relationship in terms of the geometry of the image of the absolute conic (IAC). However, it relaxes some of the constraints found in previous calibration algorithms such as [23], which requires four views to obtain four intrinsic parameters from a pair of 3D stationary points and their cast shadows. Our method is based on exploiting some novel inter-image constraints on the camera intrinsic parameters. We show that two corresponding image projections of the same 3D line on the ground plane provide two additional constraints on the camera intrinsic parameters. Consequently, two views are enough to determine the aspect ratio, the focal length, and the principal point.

Shadows are important in our perception of the world [32], because they provide important visual cues for depth and shape. They are also useful for computer vision applications such as building detection [24], surveillance systems [16,17,29,30,34,39] and inverse lighting [31,37]. However, shadows are not frequently exploited in 3D vision. Some few examples include the work on the relationship between the shadows and object structure by Kriegman and Belhumeur [22], the work on object recognition by Van Gool et al. [45], and the work on weak structured lighting by Bouguet and Perona [3]. Shadows are interesting for camera calibration since the shadows of two parallel lines cast on a planar surface by an infinite point light source (e.g. the sun), are also parallel, which together with the vertical vanishing point provide orthogonal constraints on the image of the absolute conic.

This paper extends the work in [4] by providing a solution to a more general and flexible geometric configuration, where the requirement for vertical objects is eliminated, with full performance evaluation and the analysis of degenerate cases. Our geometric constructions are close to the ones in [36] and [2]. However, our work is different from Reid and North's work [36] in three respects: First, we estimate the camera parameters and the light source orientation, while they focus on the problem of affine reconstruction of a ball out of the ground plane. Second, their main input is the image of the ground plane's horizon line, whereas in our case we assume instead the knowledge of the vertical vanishing point. Third, our major contribution is the inter-image constraints on the IAC, which lead to the use of only two images. Although Antone and Bosse [2] used shadows for camera calibration and a setup similar to ours, they require world-to-camera correspondences. Different from their techniques, the contribution of our method lies on relaxing the requirement of knowing 3D quantities, and on calibrating the camera from the information determined only from the images.

The remainder of this paper begins with an introduction to the theoretical background in Section 2. We then describe our calibration method in Section 3, followed by implementation details in Section 4, and the discussion of the degenerate configurations in Section 5. Finally, in Section 6 we present our results for both synthetic data and real images, and demonstrate applications to image-based rendering.

2. Theoretical background

2.1. Pinhole camera model and the image of the absolute conic

A 3D point $\mathbf{M} = [X\ Y\ Z\ 1]^{\mathrm{T}}$ and its corresponding image projection $\mathbf{m} = [u\ v\ 1]^{\mathrm{T}}$ are related via a 3×4 matrix $\mathbf{P}$ by

$$\mathbf{m} \sim \underbrace{\mathbf{K}[\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3\mathbf{t}]}_{\mathbf{P}}\mathbf{M}, \mathbf{K} = \begin{bmatrix} f & \gamma & u_0 \\ 0 & \lambda f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where $\sim$ indicates equality up to multiplication by a non-zero scale factor, $\mathbf{r}_i$ are the columns of the rotation matrix $\mathbf{R}$, $\mathbf{t}$ is the translation vector, and $\mathbf{K}$ is a non-singular 3×3 upper triangular matrix known as the camera calibration matrix including five parameters, i.e. the focal length f,

the skew γ, the aspect ratio λ and the principal point at (u_0, v_0).

The IAC, denoted by $\boldsymbol{\omega}$, is an imaginary point conic directly related to the camera internal matrix $\mathbf{K}$ in Eq. (1) via $\boldsymbol{\omega} = \mathbf{K}^{-\mathrm{T}}\mathbf{K}^{-1}$, which can be expanded up to a non-zero scale as

$$\boldsymbol{\omega} \sim \begin{bmatrix} 1 & -\frac{\gamma}{f\lambda} & \frac{\gamma v_0 - \lambda f u_0}{f\lambda} \\ * & \frac{f^2+\gamma^2}{f^2\lambda^2} & -\frac{\gamma^2 v_0 - \gamma\lambda f u_0 + v_0 f^2}{f^2\lambda^2} \\ * & * & \frac{v_0^2(f^2+\gamma^2) - 2\gamma v_0 \lambda f u_0}{f^2\lambda^2} + f^2 + u_0^2 \end{bmatrix}, \tag{2}$$

where the lower triangular elements are replaced by "*" to save space, since $\boldsymbol{\omega}$ is symmetric. Instead of directly determining $\mathbf{K}$, it is possible to compute the symmetric matrix $\boldsymbol{\omega}$ or its inverse (the dual image of the absolute conic), and then compute the calibration matrix uniquely using either Cholesky decomposition [35] or the following set of equations as described in [4,50]:

$$\begin{aligned} \lambda &= \sqrt{1/(\omega_{22} - \omega_{12}^2)}, \\ v_0 &= (\omega_{12}\omega_{13} - \omega_{23})/(\omega_{22} - \omega_{12}^2), \\ u_0 &= -(v_0\omega_{12} + \omega_{13}), \\ f &= \sqrt{\omega_{33} - \omega_{13}^2 - v_0(\omega_{12}\omega_{13} - \omega_{23})}, \\ \gamma &= -f\lambda\omega_{12}, \end{aligned} \tag{3}$$

where ω_{ij} denotes the element in ith row and jth column of $\boldsymbol{\omega}$.

2.2. Calibration from vanishing points

In [23], Liebowitz and Zisserman formulated the calibration constraints provided by vanishing points of mutually orthogonal directions in terms of the geometry of $\boldsymbol{\omega}$. The orthogonality relations provide linear constraints in terms of the elements of $\boldsymbol{\omega}$, which can be readily used to calibrate a camera [8,23,40,46,50]. The earlier result reported by Caprile and Torre in [6] is a special case where the camera has zero skew and unit aspect ratio. Zhang's flexible calibration method [50] uses the orthogonal property of the columns $\mathbf{r}_1$ and $\mathbf{r}_2$ of the rotation matrix $\mathbf{R}$ in Eq. (1), i.e. $\mathbf{p}_1^{\mathrm{T}}\boldsymbol{\omega}\mathbf{p}_2 = 0$, where $\mathbf{p}_i$ ($\mathbf{h}_i$ in [50], $i = 1, 2$) denotes the ith column of the projection matrix $\mathbf{P}$. This is actually equivalent to the constraint $\mathbf{v}_x^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_y = 0$, where $\mathbf{v}_x$ and $\mathbf{v}_y$ are the vanishing points along the world X- and Y-axis, respectively, because $\mathbf{p}_1$ is the scaled vanishing point $\mathbf{v}_x$ of the ray along the direction $[1\,0\,0]^{\mathrm{T}}$, and similarly $\mathbf{p}_2$ is the scaled vanishing point $\mathbf{v}_y$ of the ray along the direction $[0\,1\,0]^{\mathrm{T}}$. It is important to mention that the above derivation does not apply to the normal property of $\mathbf{r}_1$ and $\mathbf{r}_2$, i.e. $\mathbf{p}_1^{\mathrm{T}}\boldsymbol{\omega}\mathbf{p}_1 - \mathbf{p}_2^{\mathrm{T}}\boldsymbol{\omega}\mathbf{p}_2 = 0$ used in [40,50], since in general $\mathbf{v}_x$ and $\mathbf{v}_y$ have different scales, which the authors in [40,50] determine by assuming known world-to-image homographies. In [8,46], the pole–polar relationship with respect to $\boldsymbol{\omega}$ is exploited, which states that the vanishing line $\mathbf{l}$ of a plane is related to the vanishing point $\mathbf{v}$ of the normal direction to the plane via $\mathbf{l} = \boldsymbol{\omega}\mathbf{v}$. This is because any vanishing point $\mathbf{x}$ that is orthogonal to $\mathbf{v}$ must satisfy $\mathbf{x}^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v} = 0$. The important point now is that this is true for any point $\mathbf{x}$ satisfying $\mathbf{x}^{\mathrm{T}}\mathbf{l} = 0$, from which it follows that $\mathbf{l} = \boldsymbol{\omega}\mathbf{v}$.

These intra-image orthogonal constraints, together with the novel inter-image constraints that we refer to as the "*weak pole–polar*" relationship introduced in this paper, will be used later in Section 3 to derive a simple technique for camera calibration from shadows. *The weak pole–polar relationship here means that the two independent constraints on $\boldsymbol{\omega}$ arising from the pole–polar relationship can not be identified from a single view.*

3. The method

In this section, we first examine the geometry of the scene and then introduce novel inter-image constraints associated with the *weak pole–polar* relationship for camera calibration. Finally, we show how to compute the orientation of the light source.

3.1. Scene geometry

The basic geometry of a scene containing two stationary points and their shadows on the ground plane cast by an infinite point light source is shown in Fig. 2. Note that this figure shows the image projections of the 3D world points denoted by corresponding lower-case characters. For example, the world point $\mathbf{T}_1$ (not shown in Fig. 2) is mapped to $\mathbf{t}_1$ in the image plane.

It is important to point out here that our configuration shown in Fig. 2 is different from [2,4,36], which required two vertical objects and their shadows. Although, similar to [2,4,36], we assume known vertical vanishing point $\mathbf{v}_z$ and known ground plane inter-image homography $\mathbf{H}_{\boldsymbol{\pi}}$, our configuration does not assume the use of vertical objects. In other words as shown in Fig. 2 any pair of object points in 3D space and their shadows would be sufficient.

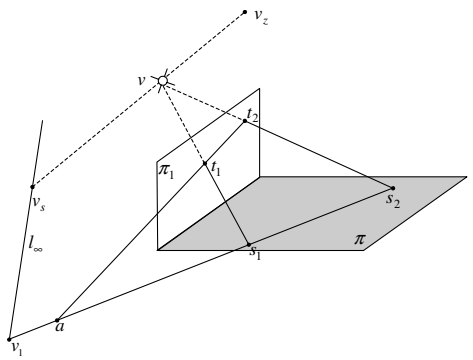

Fig. 2. Basic geometry of a scene with two stationary points $\mathbf{t}_1$ and $\mathbf{t}_2$ casting shadows $\mathbf{s}_1$ and $\mathbf{s}_2$ on the ground plane $\boldsymbol{\pi}$ by the distant light source $\mathbf{v}$, i.e. the sun. $\boldsymbol{\pi}_1$ is the plane perpendicular to the ground plane that is passing through the two points $\mathbf{t}_1$ and $\mathbf{t}_2$. $\mathbf{l}_\infty$ is the horizon line of the ground plane $\boldsymbol{\pi}$, where the any parallel lines in the ground plane through the two shadow points would meet, e.g. the point $\mathbf{v}_s$.

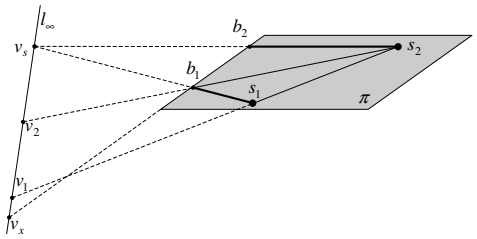

Fig. 3. The inter-image constraints in Eq. (12) can be applied to any corresponding ground lines between two views (except for $\mathbf{b}_i\mathbf{s}_i$). For example, we have the image correspondences, $\mathbf{b}_1\mathbf{b}_2$ and $\mathbf{b}'_1\mathbf{b}'_2$, of the 3D line $\mathbf{B}_1\mathbf{B}_2$ between two views, and thus have the inter-image constraints on the vanishing point $\mathbf{v}_x$, which is the intersection of the line $\mathbf{b}_1\mathbf{b}_2$ and $\mathbf{l}_\infty$.

ficient to perform calibration from two views. Let us now denote the orthogonal projections of the 3D points $\mathbf{t}_1$ and $\mathbf{t}_2$ in the ground plane as $\mathbf{b}_1$ and $\mathbf{b}_2$, which we shall hereafter refer to as the footprints of the 3D points. In [2,4,36], the assumption is that the footprints are available and hence the vanishing point $\mathbf{v}_s$ as shown in Fig. 3 can be directly identified in each image. In our case, since we do not limit ourselves to vertical objects, i.e. $\mathbf{b}_1$ and $\mathbf{b}_2$ are unknown, the vanishing point $\mathbf{v}_s$ can not be directly identified in each image. This new configuration is of course more practical since even when we deal with vertical objects it is not always trivial to detect the footprints due to occlusion or the thickness of the vertical objects, while it is typically possible to compute vertical vanishing point $\mathbf{v}_z$ from all available vertical directions in the image. Moreover, shadows cast by not necessarily vertical objects can also be used for calibration, in which case the vanishing point $\mathbf{v}_s$ is not available.

In this setup, we can also compute the imaged sun's position $\mathbf{v}$ as

$$\mathbf{v} \sim (\mathbf{t}_1 \times \mathbf{s}_1) \times (\mathbf{t}_2 \times \mathbf{s}_2). \tag{4}$$

Clearly, any two parallel lines in the ground plane through the two shadow points $\mathbf{s}_1$ and $\mathbf{s}_2$ would meet on the vanishing line of the ground plane. However, if in addition the two parallel lines pass through the footprints $\mathbf{b}_1$ and $\mathbf{b}_2$, then their intersection point $\mathbf{v}_s$ on the vanishing line would be collinear with $\mathbf{v}_z$, and $\mathbf{v}$. This collinearity allows us to recover $\mathbf{v}_s$ even without the knowledge of the footprints as shown below. The planes passing through $(\mathbf{v}_z, \mathbf{t}_2, \mathbf{s}_2)$ and $(\mathbf{v}_z, \mathbf{t}_1, \mathbf{s}_1)$ are parallel and intersect at the vanishing line $\mathbf{v}_z \times \mathbf{v}$. On the other hand, the vanishing line $\mathbf{v}_z \times \mathbf{v}$ intersects the ground plane at $\mathbf{v}_s$. Therefore, the point $\mathbf{v}_s$ in the first image can be computed as

$$\mathbf{v}_s = (\mathbf{H}_\pi^{\mathrm{T}}(\mathbf{v}'_z \times \mathbf{v}')) \times (\mathbf{v}_z \times \mathbf{v}), \tag{5}$$

where all primed variables indicate the corresponding points in the second image. Of course, $\mathbf{v}'_s$ in the second image is then readily computed as $\mathbf{H}_\pi\,\mathbf{v}_s$, and hence the footprints $\mathbf{b}_i$ can be recovered using

$$\mathbf{b}_i = (\mathbf{v}_s \times \mathbf{s}_i) \times (\mathbf{v}_z \times \mathbf{t}_i). \tag{6}$$

Our configuration now reduces to the one used in [2,4,36].

Since the vertical vanishing point $\mathbf{v}_z$ and the horizon line $\mathbf{l}_\infty$ of the ground plane $\boldsymbol{\pi}$ have a pole-polar relationship with respect to the image of the absolute conic $\boldsymbol{\omega}$, we have

$$\mathbf{v}_1 \times \mathbf{v}_s \sim \mathbf{l}_\infty \sim \boldsymbol{\omega}\mathbf{v}_z, \tag{7}$$

where $\mathbf{v}_1$ is the intersection of the line $\mathbf{s}_1\,\mathbf{s}_2$ and the horizon line $\mathbf{l}_\infty$, i.e. another vanishing point on the ground plane. Eq. (7) can be rewritten, equivalently, as two constraints on $\boldsymbol{\omega}$:

$$\mathbf{v}_s^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_z = 0, \tag{8}$$

$$\mathbf{v}_1^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_z = 0. \tag{9}$$

In our case, we only have the constraint in Eq. (8) since we can not determine $\mathbf{v}_1$ from a single view. Without further assumptions, it is unlikely that we can extract more constraints on the camera internal parameters from a single view of such a scene as shown in Fig. 2. Therefore, we use two images and introduce the *weak pole–polar* constraint on the IAC to solve the camera calibration.

Before we move further, we mention some possible configurations that may provide more constraints, although we will not make use of such constraints. One possibility is to use the ratio between the heights h_1 and h_2 of the two points $\mathbf{t}_1$ and $\mathbf{t}_2$ from the ground plane since

$$\frac{h_1}{h_2} = \frac{(\mathbf{a} - \mathbf{s}_1)(\mathbf{v}_1 - \mathbf{s}_2)}{(\mathbf{a} - \mathbf{s}_2)(\mathbf{v}_1 - \mathbf{s}_1)}, \tag{10}$$

\where $\mathbf{a} \sim (\mathbf{t}_1 \times \mathbf{t}_2) \times (\mathbf{s}_1 \times \mathbf{s}_2)$. Eq. (10) can be directly used to recover the second vanishing point $\mathbf{v}_1$ on the horizon line $\mathbf{l}_\infty$. One special case is that $\mathbf{t}_1$ and $\mathbf{t}_2$ have the same heights, in which case the vanishing point $\mathbf{v}_1$ coincides with the intersection $\mathbf{a}$. Other possibilities include using the knowledge of the orientation of the light source $\mathbf{v}$ as shown in [2]. All these cases provide more constraints than the one we have in Eq. (8), and therefore it is possible to calibrate the camera using one view by assuming a simplified camera model [6,23]. However, we keep our problem more general to extend the applicability of the solution.

3.2. Inter-image constraints

The second view can be easily used to obtain another constraint $\mathbf{v}_s^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_z = 0$ from Eq. (8). Beyond that, we explore here the inter-image constraints associated with the *weak pole–polar* relationship as described below. Geometrically, $\mathbf{v}_1^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_z = 0$ in Eq. (9) can be interpreted as the statement that the vanishing point $\mathbf{v}_1$ of the line $\mathbf{s}_1\mathbf{s}_2$ lies on the horizon line $\boldsymbol{\omega}\mathbf{v}_z$ of the ground plane. Considering also that $\mathbf{v}_1$ lies on the imaged line $\mathbf{s}_1\mathbf{s}_2$, we can express $\mathbf{v}_1$ as a function of $\boldsymbol{\omega}$:

$$\mathbf{v}_1 \sim [\mathbf{s}_1 \times \mathbf{s}_2]_\times \boldsymbol{\omega}\mathbf{v}_z, \tag{11}$$

where $[\cdot]_\times$ is the usual notation for the skew symmetric matrix characterizing the cross product. We call this relationship the *weak pole–polar* relationship in the sense that we can not identify the polar, $\mathbf{l}_\infty = \boldsymbol{\omega}\mathbf{v}_z$, to the pole $\mathbf{v}_z$ with

respect to $\boldsymbol{\omega}$ from a single view, since we only have one constraint $\mathbf{v}_s^{\mathrm{T}}\mathbf{l}_\infty = 0$ on $\mathbf{l}_\infty$.

The *weak pole–polar* relationship provides inter-image constraints on the image of the absolute conic, and hence can be exploited for calibration. For this purpose, note that the ground plane homography $\mathbf{H}_\pi$ and the fundamental matrix $\mathbf{F}$, which define the inter-image relations, can be recovered in practice by a set of point correspondences. We then have the following inter-image constraints in terms of $\boldsymbol{\omega}$,

$$\mathbf{v}_1' \sim \mathbf{H}_\pi \mathbf{v}_1, \text{ and } \mathbf{v}_1'^{\mathrm{T}}\mathbf{F}\mathbf{v}_1 = 0, \tag{12}$$

where $\mathbf{v}_1'$ is the corresponding vanishing point of $\mathbf{v}_1$ in the second image that can also be expressed as a function of $\boldsymbol{\omega}$ using Eq. (11).

One interesting observation is that the inter-image constraints in Eq. (12) can also be applied to any corresponding lines on the ground plane between two views, except for $\mathbf{s}_1\mathbf{b}_1$ and $\mathbf{s}_2\mathbf{b}_2$ in Fig. 3, because the concurrency of all such lines at the point $\mathbf{v}_s$ on the ground plane has already been enforced in Eq. (8). For example, we have the image correspondences, $\mathbf{b}_1\mathbf{b}_2$ and $\mathbf{b}_1'\mathbf{b}_2'$, of the 3D line $\mathbf{B}_1\mathbf{B}_2$ between two views, and thus have the inter-image constraints on the vanishing point $\mathbf{v}_x$, which as shown in Fig. 3 is the intersection of the line $\mathbf{b}_1\mathbf{b}_2$ and $\mathbf{l}_\infty$

$$\mathbf{v}_x \sim [\mathbf{b}_1 \times \mathbf{b}_2]_\times \boldsymbol{\omega}\mathbf{v}_z. \tag{13}$$

Consequently, we can obtain the correct solution for $\boldsymbol{\omega}$ by minimizing the following symmetric transfer errors of the geometric distances and epipolar distances of the vanishing points on the line $\mathbf{l}_\infty$, which are in terms of the elements of $\boldsymbol{\omega}$:

$$\sum_i \{d_1(\mathbf{v}_i', \mathbf{H}_\pi \mathbf{v}_i)^2 + d_2(\mathbf{v}_i', \mathbf{F}\mathbf{v}_i)^2\}, \tag{14}$$

where $d_1(\mathbf{x},\mathbf{y})$ is the geometric image distance between the two homogeneous image points represented by $\mathbf{x}$ and $\mathbf{y}$, and $d_2(\mathbf{x},\mathbf{l})$ is the geometric image distance from an image point $\mathbf{x}$ to an image line $\mathbf{l}$.

Generally, the inter-image relations in (14) provide two constraints on $\boldsymbol{\omega}$. As a result, under the assumption of fixed intrinsic parameters and zero skew, it is possible to calibrate a camera from two or more views of a scene containing two stationary points and their cast shadows. The assumption of zero skew is reasonable both theoretically and practically. As pointed out in [14], the skew will be zero for most normal cameras and can take non-zero values only in certain unusual instances, e.g. taking an image of an image. This argument also confirms some previous observations, e.g. [4,23,40,51].

Different from the linear orthogonal constraints in Eq. (8), the inter-image constraints in Eq. (14) are not linear in the elements of $\boldsymbol{\omega}$, and hence need a non-linear optimization process. The starting point can be obtained as follows. As shown in Fig. 2, the four points $\mathbf{v}_1$, $\mathbf{s}_2$, $\mathbf{s}_1$, and $\mathbf{a}$, are collinear. Therefore, their cross-ratio is preserved under perspective projection, i.e.

$$\{\mathbf{v}_1, \mathbf{s}_2; \mathbf{s}_1, \mathbf{a}\}^1 = \{\mathbf{v}_1, \mathbf{s}_2; \mathbf{s}_1, \mathbf{a}\}^2, \tag{15}$$

where $\{\cdot,\cdot;\cdot,\cdot\}$ denotes the cross-ratio of four points, and the superscripts indicate the images in which the cross-ratios are taken. This provides a third constraint on $\boldsymbol{\omega}$, which can be defined by a 6D vector with five unknowns:

$$\mathbf{w} \sim [1, \omega_{12}, \omega_{22}, \omega_{13}, \omega_{23}, \omega_{33}]^{\mathrm{T}}. \tag{16}$$

Assuming that the camera has zero skew and unit aspect ratio (resulting in $\omega_{12} = 0$ and $\omega_{22} = 1$), these three known constraints (two from the orthogonality constraints $\mathbf{v}_s^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_z = 0$ in (8) and one from (15)) are sufficient to solve for the three unknowns: ω_{13}, ω_{23} and ω_{33}. Note that, (15) is quadratic in terms of the elements of $\boldsymbol{\omega}$, leading to a two-fold ambiguity in the starting point. This ambiguity is eliminated during the minimization of the cost function in (14), leading to the correct solution.

3.3. Estimation of light source orientation

After calibrating the camera, we have no difficulty in estimating the light source position and orientation provided that the corresponding feature points along the lighting direction can be identified from two views. For example, we can compute the orientation of the light source in 3D by using the optimal triangulation method [13] as follows. Without loss of generality, we choose the world coordinate frame of the scene shown in Fig. 2 as follows: origin at $\mathbf{B}_2$ (which is not visible but whose projection can be computed using Eq. (6)), Z-axis along the line $\mathbf{B}_2$ $\mathbf{T}_2$ with the positive direction towards $\mathbf{T}_2$, X-axis along the line $\mathbf{B}_1\mathbf{B}_2$ with the negative direction towards $\mathbf{B}_1$, and the Y-axis given by the right-hand rule. Then, we use the triangulation method to compute the 3D locations of $\mathbf{B}_1$ and $\mathbf{S}_1$, and the orientation of the light source is then given by $\mathbf{n} = \mathbf{B}_1 - \mathbf{S}_1$.

Alternatively, since in our case the light source is far away, we only need to measure the orientation of the light source, which can be expressed by two angles: the polar angle ϕ between the lighting direction and the vertical Z-axis, and the azimuthal angle θ in the ground plane from the X-axis (the line $\mathbf{B}_1\mathbf{B}_2$). Note that the imaged light source $\mathbf{v}$ is the vanishing point along lighting direction since the light source is on $\boldsymbol{\pi}_\infty$, and $\mathbf{v}_s$ is the projection of $\mathbf{v}$ on the ground plane. Consequently, ϕ and θ can be measured from the corresponding vanishing points using

$$\phi = \cos^{-1} \frac{\mathbf{v}_z^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}}{\sqrt{\mathbf{v}^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}}\sqrt{\mathbf{v}_z^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_z}}, \tag{17}$$

$$\theta = \cos^{-1} \frac{\mathbf{v}_x^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_s}{\sqrt{\mathbf{v}_s^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_s}\sqrt{\mathbf{v}_x^{\mathrm{T}}\boldsymbol{\omega}\mathbf{v}_x}}. \tag{18}$$

In our experiments, we used (17) and (18) to compute ϕ and θ for each view and averaged the results.

4. Implementation details

4.1. Feature extraction

Features, extracted either automatically, e.g. using a corner detector, or interactively, are subject to errors. The features are mostly the image locations $\mathbf{t}_i$ and $\mathbf{s}_i$ that are extracted directly from images. In our implementation, we use a maximum likelihood estimation method similar to [9] with the uncertainties in the locations of points $\mathbf{t}_i$ and $\mathbf{s}_i$ modeled by the covariance matrices $\mathbf{\Lambda}_{\mathbf{t}_i}$ and $\mathbf{\Lambda}_{\mathbf{s}_i}$ respectively.

In the error-free case, imaged shadow relations (illuminated by a point light source) are modeled by a *planar homology* [14] (see Fig. 4 and also Fig. 5). The light source has to be a point light source, i.e. all light rays are concurrent. A planar homology is a special case of a planar homography which has a line of fixed points, called the *axis*, and a distinct fixed point $\mathbf{v}$, not on the line, called the *vertex* of the homology. In our case, the axis is the line $\mathbf{b}_1\mathbf{b}_2$ and the vertex $\mathbf{v}$ is the image of the light source, located outside the image frames in Fig. 5 but illustrated in Fig. 4. Under this transformation, points on the axis are mapped to themselves. Each point off the axis, e.g. $\mathbf{t}_2$, lies on a fixed line $\mathbf{t}_2\mathbf{s}_2$ through $\mathbf{v}$ intersecting the axis at $\mathbf{i}_2$ and is mapped to another point $\mathbf{s}_2$ on the line. Note that $\mathbf{i}_2$ is the intersection in the image plane, although the light ray $\mathbf{t}_2\mathbf{s}_2$ and the axis, $\mathbf{b}_1\mathbf{b}_2$, are unlikely to intersect in the 3D world. Consequently, corresponding point pairs, $\mathbf{t}_i \leftrightarrow \mathbf{s}_i$, and the vertex $\mathbf{v}$ are collinear.

One important property of a planar homology is that the corresponding lines, i.e. lines through pairs of corresponding points, intersect on the axis. For example, the lines $\mathbf{t}_1\mathbf{t}_2$ and $\mathbf{s}_1\mathbf{s}_2$ satisfy

$$\mathbf{b}_1\mathbf{b}_2 \cdot (\mathbf{t}_1\mathbf{t}_2 \times \mathbf{s}_1\mathbf{s}_2) = 0. \tag{19}$$

Another important property of a planar homology is that the cross ratio defined by the vertex, $\mathbf{v}$, the corresponding points, $\mathbf{t}_i$ and $\mathbf{s}_i$, and the intersection, $\mathbf{i}_i$, of their join $\mathbf{t}_i\mathbf{s}_i$ with the axis, is the characteristic invariant of the homology, and is the same for all corresponding points. For example, the cross-ratios of the four points $\{\mathbf{v}, \mathbf{t}_1; \mathbf{s}_1, \mathbf{i}_1\}$ and $\{\mathbf{v}, \mathbf{t}_2; \mathbf{s}_2, \mathbf{i}_2\}$ are equal,

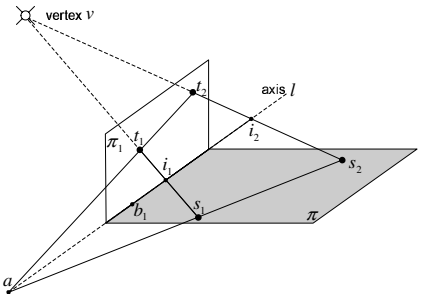

Fig. 4. Geometrically, a planar object, π_1, and its shadow (cast on a ground plane π) are related by a planar homology.

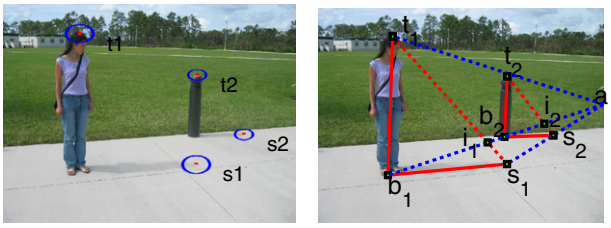

Fig. 5. Maximum likelihood estimation of the feature points: (Left) The uncertainty ellipses are shown. These ellipses are specified by the user, and indicate the confidence region for localizing the points. (Right) Estimated features points are satisfying the alignment constraints in Eq. (19), (21) and the invariance constraint in Eq. (20).

$$\{\mathbf{v}, \mathbf{t}_1; \mathbf{s}_1, \mathbf{i}_1\} = \{\mathbf{v}, \mathbf{t}_2; \mathbf{s}_2, \mathbf{i}_2\}. \tag{20}$$

In addition, the four points $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{b}_1$, and $\mathbf{b}_2$ should lie on the ground plane,

$$\mathbf{x}'_j = \mathbf{H}_\pi \mathbf{x}_j, \text{ where } j = 1, 2 \text{ and } \mathbf{x} = \mathbf{t}, \mathbf{s}. \tag{21}$$

Therefore, we determine the maximum likelihood estimates of the true locations of the features by minimizing the following sum of Mahalanobis distances between the input points and their MLE estimates subject to the alignment constraints in (19), (21) and the invariant constraint in (20):

$$\arg\min_{\hat{\mathbf{x}}_j} \sum_{j=1,2} \sum_{\mathbf{x}=\mathbf{t},\mathbf{s}} (\mathbf{x}_j - \hat{\mathbf{x}}_j)^{\mathrm{T}} \mathbf{\Lambda}_{\mathbf{x}_j}^{-1} (\mathbf{x}_j - \hat{\mathbf{x}}_j). \tag{22}$$

The covariance matrices $\mathbf{\Lambda}_{\mathbf{t}_i}$ and $\mathbf{\Lambda}_{\mathbf{s}_i}$ are not necessarily isotropic or equal. For example, in our experiments for the image shown in Fig. 5, the second view of the first real image set shown in Fig. 9, we set $\mathbf{\Lambda}_{\mathbf{t}_1} = \mathrm{diag}([20^2, 10^2])$ and $\mathbf{\Lambda}_{\mathbf{t}_2} = \mathbf{\Lambda}_{\mathbf{s}_1} = \mathbf{\Lambda}_{\mathbf{s}_1} = 8^2\mathbf{I}_{2\times 2}$. The image is of size 2272×1704. In our implementation, we assumed the uncertainties in the locations of points $\mathbf{t}_i$ and $\mathbf{s}_i$ are similar between views, and hence we used the same $\mathbf{\Lambda}_{\mathbf{t}_i}$ and $\mathbf{\Lambda}_{\mathbf{s}_i}$ for all images. Note that, in order to obtain more robust solutions, we also computed point correspondences other than the four feature points $\mathbf{t}_i$ and $\mathbf{s}_i$, $i = 1, 2$, for the computation of inter-image planar homographies and the fundamental matrix in Eq. (14).

4.2. Algorithm outline

The complete algorithm will now be summarized. In addition to the vertical vanishing point and the inter-image planar homography induced by the ground plane, the input includes a pair of images of a scene containing two stationary points and their shadows on the ground plane cast by an infinite light (e.g. the sun). The footprints of the two stationary points are not necessarily visible in any of the images. The outputs are the camera parameters and the orientation of the light source. A top-level outline of the algorithm is as follows:

(1) Identify a seed set of image-to-image matches between the two images. At least four points, $\hat{\mathbf{t}}_i$ and $\hat{\mathbf{s}}_i$ $(i = 1, 2)$,

are needed, though more are preferable. The feature points can be refined as shown in Section 4.1.

(2) Obtain the initial closed form solution
   (a) Compute the sun position $\mathbf{v}$ using (4) and shadow vanishing point $\mathbf{v}_s$ using (5) for each view, and thus obtain two linear constraints on $\boldsymbol{\omega}$ using (8).
   (b) Compute the third constraint using (15).
   (c) Solve the two linear constraints in (8) and one quadratic constraint in (15) by assuming $\omega_{12} = 0$ and $\omega_{22} = 1$. This gives us ω_{13}, ω_{23} and ω_{33} up to an ambiguity.

(3) Obtain the refined solution
   (a) Compute the fundamental matrix $\mathbf{F}$ (when enough point correspondences are available) as described in Section 3.2.
   (b) Minimize the cost function in (14) using the above ω_{13}, ω_{23}, and ω_{33} as starting points. The ambiguity in (c) is eliminated during the minimization process, leading to the correct solution.

(4) Use bundle adjustment [43] to refine further the solution.

(5) Compute the orientation of the light source using (17) and (18).

5. Degenerate configurations

The proposed algorithm, like almost any algorithm, has degenerate configurations for which the parameters can not be estimated. In practice, it is important to be aware of such singularities in order to obtain reliable results by avoiding them.

Generally, as an approach based on vanishing points, our method degenerates if the vanishing points along some directions are at infinity. Geometrically, vanishing points go to infinity when the perspective transformation degenerates to affine transformation, in which case parallel lines in the 3D space remain also parallel in the image plane. This occurs when the direction of the vanishing point is parallel to the image plane, since the vanishing point of a line is obtained by intersecting the image plane with a ray parallel to the world line and passing through the camera center. Algebraically, this means that the third component of the homogeneous vanishing point coordinates becomes equal to zero. In practice, if conditions are near degenerate, then the solution is ill-conditioned, and the particular member of the family of solutions is determined by "noise". Therefore, for calibration algorithms based on vanishing points, when possible, it is important to work with images that have large perspective distortions.

In our paper, we mainly employ the orthogonality relationship in (8) of the vanishing points, $\mathbf{v}_z$ and $\mathbf{v}_s$, and the inter-image constraints (14) and (15) associated with vanishing points other than $\mathbf{v}_s$ on the horizon line $\mathbf{l}_\infty$. Therefore, we are aware of the vanishing point, $\mathbf{v}_z$, along the vertical direction, and other vanishing points of world lines that are perpendicular to the vertical direction, i.e. vanishing points on the line $\mathbf{l}_\infty$, as shown in Fig. 2. We first derive the singularities of our method with geometric explanations. Then, we give algebraic analysis on the situations where the inter-image constraints fail. For this analysis, we use the vanishing point $\mathbf{v}_1$ (Eq. (11)) as an example, but the derivation can be easily extended to others.

5.1. Vanishing point at infinity

Liebowitz and Zisserman [23] demonstrate an easy but efficient method to identify the degeneracies that occur when linear constraints on $\boldsymbol{\omega}$ are not independent. However, our calibration method introduces some nonlinear constraints in (14) and (15), which make it more difficult to identify degenerate configurations. Although it is still possible to apply the approach proposed in [23] to the linear constraints in (8), we derive all possible singularities with geometrical explanations.

The following analyses are mostly related to [40] by Sturm and Maybank but different mainly in three respects. First, although both their and our approaches use the orthogonality constraints in Eq. (8), they have additional constraints arising from the normal properties of $\mathbf{r}_1$ and $\mathbf{r}_2$ as described in Section 2.2, while we derive new inter-image constraints in (14) and (15). Notice that in their case the use of normal properties of $\mathbf{r}_1$ and $\mathbf{r}_2$ is viable because they have the world to image planar homography. Second, the two planes $\boldsymbol{\pi}$ and $\boldsymbol{\pi}_1$ (See Fig. 2) in our case are perpendicular to each other, which is a special case of the configurations addressed in [40]. Third, only a subset of their degenerate cases apply to our setup.

Our method degenerates in the following cases: (i) the image plane is parallel to the plane $\boldsymbol{\pi}_1$, (ii) the image plane is parallel to the ground plane $\boldsymbol{\pi}$, and (iii) the image plane is perpendicular to both $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}$. In case (i), both $\mathbf{v}_x$ and $\mathbf{v}_z$ go to infinity. In case (ii), all the vanishing points on the line $\mathbf{l}_\infty$ will be infinite. The case (ii) is different from case (i) in that the principal point can be recovered in case (ii), but not in case (i). In case (ii), the vanishing points on the ground plane $\boldsymbol{\pi}$, e.g. $\mathbf{v}_x$ and $\mathbf{v}_s$, are the vanishing points for directions parallel to the image plane. Therefore, the vertical direction, which is perpendicular to the ground plane $\boldsymbol{\pi}$, must be orthogonal to the image plane, and parallel to the principal axis of the camera. The vanishing point of this principal axis is thus the principal point, which is given by the finite vanishing point $\mathbf{v}_z$. The same derivation is valid for case (i) only if $\mathbf{v}_s$ is orthogonal to $\mathbf{v}_x$, which is not necessarily true. Note that the result in [40] that the aspect ratio λ can be recovered when one plane is parallel to the image plane does not apply to our case, since we have no Euclidean information about the world coordinates. In case (iii), if the degenerate situation happens for both views, we are able to intersect the two lines $\mathbf{b}_1\mathbf{b}_2$ in two views to obtain the principal point as analyzed above. Partial analyses are validated in Section 6.1.

5.2. Difficulty in computing $\mathbf{v}_1$

One important step of the proposed algorithm is to express $\mathbf{v}_1$ as a function of $\boldsymbol{\omega}$ in Eq. (11). Therefore, our method degenerates when $\mathbf{v}_1$ can not be computed from Eq. (11). In other words, when the lines $\mathbf{s}_1$ $\mathbf{s}_2$ and $\mathbf{l}_\infty$ are parallel to each other in the image plane. In order to simplify the analysis, we assume that the world origin is at $\mathbf{S}_2$, Z-axis is along the line $\mathbf{B}_2$ $\mathbf{T}_2$ with the positive direction towards $\mathbf{T}_2$, X-axis is along the line $\mathbf{S}_1\mathbf{S}_2$ with the negative direction towards $\mathbf{S}_1$, and the Y-axis is given by the right-hand rule. Note that the algebraic derivation is independent of the world coordinate definition.

The line $\mathbf{s}_1$ $\mathbf{s}_2$ is the imaged projection of the 3D X-axis, and thus passes through the vanishing point $\mathbf{v}_1$ ($\sim\mathbf{p}_1$) along the direction $[1\,0\,0]^{\mathrm{T}}$ and the image of the world origin given by

$$\mathbf{s}_2 \sim [\mathbf{p}_1\ \mathbf{p}_2\ \mathbf{p}_3\ \mathbf{p}_4][0\ 0\ 0\ 1]^{\mathrm{T}} = \mathbf{p}_4. \tag{23}$$

Consequently, it can be expressed as

$$\mathbf{s}_1\mathbf{s}_2 \sim \mathbf{p}_1 \times \mathbf{p}_4. \tag{24}$$

Similarly, the line $\mathbf{l}_\infty$ passes through the two vanishing points $\mathbf{v}_x$ (i.e. $\mathbf{v}_1 \sim \mathbf{p}_1$) and $\mathbf{v}_y(\sim\mathbf{p}_2)$,

$$\mathbf{l}_\infty = \mathbf{p}_1 \times \mathbf{p}_2. \tag{25}$$

When the lines $\mathbf{s}_1\mathbf{s}_2$ and $\mathbf{l}_\infty$ are parallel to each other, they intersect at infinity, i.e. the third component of the imaged intersection, $\mathbf{b}_1\mathbf{s}_2 \times \mathbf{l}_\infty$, equals to zero. By using the property that the cofactor matrix is related to the way matrices distribute with respect to the cross product [14], we have

$$\begin{aligned}
\{\mathbf{s}_1\mathbf{s}_2 \times \mathbf{l}_\infty\}_3 = 0 &\Longleftrightarrow \{(\mathbf{p}_1 \times \mathbf{p}_4) \times (\mathbf{p}_1 \times \mathbf{p}_2)\}_3 = 0 \\
&\overset{\text{Eq.(1)}}{\Longleftrightarrow} \{[(\mathbf{K}\mathbf{r}_1) \times (-\mathbf{K}\mathbf{R}\mathbf{C})] \times (\mathbf{K}\mathbf{r}_1 \times \mathbf{K}\mathbf{r}_2)\}_3 = 0 \\
&\Longleftrightarrow \{[\mathbf{K}^*(\mathbf{r}_1 \times \mathbf{R}\mathbf{C})] \times [\mathbf{K}^*(\mathbf{r}_1 \times \mathbf{r}_2)]\}_3 = 0 \\
&\Longleftrightarrow \{\mathbf{K}[(\mathbf{r}_1 \times \mathbf{R}\mathbf{C}) \times \mathbf{r}_3]\}_3 = 0 \\
&\Longleftrightarrow \{\mathbf{K}[([\mathbf{r}_1]_\times[\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3]\mathbf{C}) \times \mathbf{r}_3]\}_3 = 0 \\
&\Longleftrightarrow \{\mathbf{K}(C_y\mathbf{r}_3 - C_z\mathbf{r}_2) \times \mathbf{r}_3\}_3 = 0 \\
&\Longleftrightarrow \{C_z\mathbf{K}\mathbf{r}_1\}_3 = 0 \\
&\Longleftrightarrow C_z = 0 \text{ or } \{\mathbf{v}_x\}_3 = 0,
\end{aligned} \tag{26}$$

where $\mathbf{C} = [C_x\ C_y\ C_z]^{\mathrm{T}}$ is the camera projection center, $\mathbf{K}^*$ denotes the matrix of cofactors of $\mathbf{K}$, and $\{\mathbf{z}\}_3$ denotes the 3rd element of the vector $\mathbf{z}$. Algebraically, $C_z = 0$ implies that the line $\mathbf{b}_1\mathbf{b}_2$ coincides with the horizon line $\mathbf{l}_\infty$, in which case the pole–polar relationship in Eq. (7) offers two independent constraints on the camera internal parameters per view. The two constraints might be used in camera calibration. However, the configuration $C_z = 0$ indicates that the camera center lies on the ground plane $\boldsymbol{\pi}$ in Fig. 2, which is unlikely to happen in practice since it is rare to capture images or videos like that. Therefore, we only need to avoid the case $\{\mathbf{v}_x\}_3 = 0$, in which case $\mathbf{v}_x$ goes to infinity.

6. Experimental results

The proposed method can be directly used for applications where it is difficult to calibrate offline using a special calibration pattern with known geometry, and when the number of views available is insufficient to employ a self-calibration technique. Below, we provide thorough experiments for both simulated and real data, and show an application to image-based rendering.

6.1. Computer simulation

The simulated camera has the focal length of $f = 1000$, the aspect ratio of $\lambda = 1.06$, the skew of $\gamma = 0$, and the principal point at $u_0 = 8$ and $v_0 = 6$. The synthetic light source is at infinity with the polar angle $\phi = \arctan(0.5)$ degrees and the azimuthal angle $\theta = 60°$. The two stationary 3D points have heights of 100 and 80 U from the ground plane, respectively. The distance between the two footprints on the ground plane is 75 U. In order to approximate the real cases where there existing some other point correspondences besides the two 3D stationary points and their cast shadows, we generated ten 3D points randomly located on a hemisphere centered at $(75, 0, 0)$ with radius 80, which is above the ground plane, and another ten 3D points on the ground plane. In the experiments presented herein, we generated four views with the camera and the "at" positions listed in Table 1, where we follow camera (eye) coordinate specification in OpenGL fashion. Therefore, *at-camera* is the principal view direction.

6.1.1. Performance versus noise level

In this experimentation, we used the two combinations of image pairs (views) in Table 1. The first combination composes of 1st and 4th views, while the second one includes 2nd and 3rd views. Gaussian noise with zero mean and a standard deviation of $\sigma \leqslant 3.0$ pixels was added to the projected image points. The estimated camera parameters were then compared with the ground truth. For each noise level, we performed 1000 independent trials. The final averaged results of camera parameters and light source orientation are shown in Fig. 6. As argued by [42,51], the relative difference with respect to the focal length rather than the absolute error is a more geometrically meaningful error measure. Therefore, we measured the relative error of f, u_0, and v_0 with respect to f while varying the noise level from 0.3 pixels to 3.0 pixels. For the aspect ratio λ, we measured the relative error with respect to its ground truth. Obviously, errors increase almost linearly with respect to the noise level for both the camera internal parameters and the light source orientations. For $\sigma = 1.5$, which is comparable to the typical noise in practical calibration, the relative error of focal length f is 1.93% for the first combination and 2.57% for the second combination. For more noise, the relative error in focal length increases steadily to reach 7.16% for the first combination and 5.85% for the second one at $\sigma = 3.0$. The maximum relative error in the

Table 1
External parameters of four different viewpoints

View	1st	2nd	3rd	4th
Camera position	$(10,-100,40)$	$(-150,-100,40)$	$(10,-100,100)$	$(-150,-100,100)$
"at" position	$(-100,0,0)$	$(-100,0,0)$	$(-100,0,0)$	$(-100,0,0)$

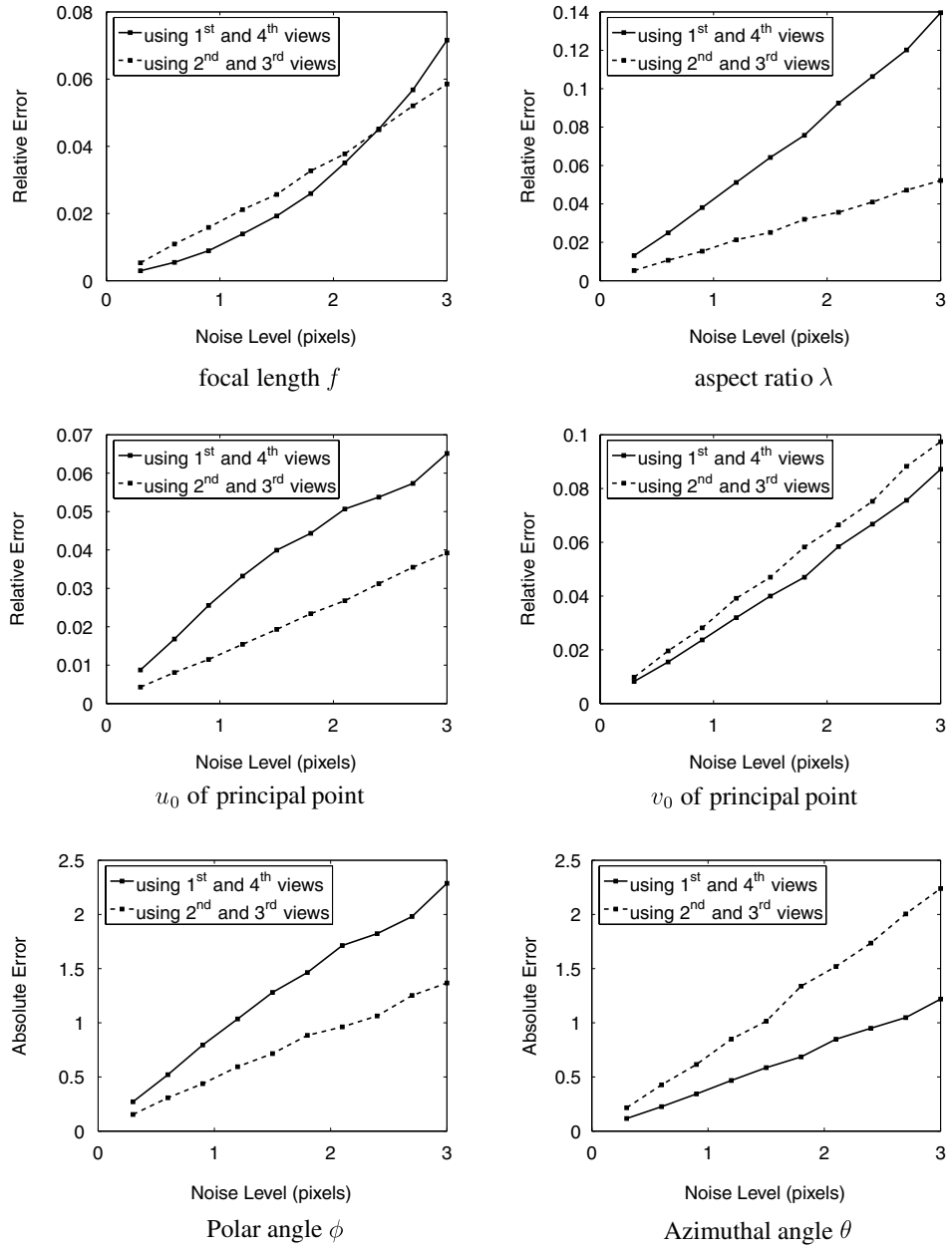

Fig. 6. The performance of our method under different noise levels (in pixels).

aspect ratio is 13.96% for the first combination, and 5.22% for the second combination. The maximum relative errors in the coordinates of the principal point are around 6.51% for u_0 and about 9.74% for v_0. Finally, the computed orientation of the light source is within $\pm 2.5^\circ$ for both polar and azimuthal angles.

6.1.2. Performance versus the orientation of the image plane

Another experiment performed was to examine the influence of the relative orientation between the vertical plane $\boldsymbol{\pi}_1$ (see Fig. 2) and the image plane. Two views were used. The orientation of the image planes of two views were chosen as follows: the image planes were initially parallel to the vertical plane; a rotation axis was randomly chosen from a uniform sphere; the image planes were then rotated around that axis with angle ψ. Considering the fact that extracted feature points will in practice be affected by noise, we also added a typical noise level of $\sigma = 1.0$ pixels to all projected image points. We repeated this process 1000 times and computed the average errors. The angle ψ is varied from 5° to 85°, and the results are shown in Fig. 7. Several important observations are made immediately. First, the results validate the partial analysis of degenerate cases in Section 5.1. In case (i) where the image plane is parallel to the vertical plane, i.e. $\psi = 0°$, our method degenerates. In case (iii), i.e. $\psi = 90°$, our method is also degenerate, but is able to recover the principal point since in our experiments the image planes of both views are also approximately perpendicular to the ground plane. The second observation is that the best performance seems to be achieved with an angle around 45°. Third, the minima of the curve is sufficiently broad so that in practice acceptable results can be found from a wide range of viewing angles. Note that, when ψ increases, foreshortening makes the feature detection less accurate, which is not considered in these experiments.

6.1.3. Performance versus orthogonality errors

This last experiment was carried out to evaluate how sensitive the algorithm is to errors in estimating the footprints $\mathbf{b}_i$, which enforce the fact that $\mathbf{t}_i\mathbf{b}_i$ must be orthogonal to the ground plane. For each direction $\mathbf{t}_i\mathbf{b}_i$, we added independent Gaussian noise with zero mean and a standard deviation of $\sigma \leqslant 3.0°$. Considering the fact that the extracted feature points will in practice be affected by noise, we also added a typical noise level of $\sigma = 1.0$ pixels to all projected image points. We repeated this process 1000 times, and the final averaged results of camera parameters and light source orientation are shown in Fig. 8. The errors in both camera parameters and light source orientation increase almost linearly with respect to the orthogonality errors. Notice that the errors do not go to zero as orthogonality errors tend towards zero due to the added noise in image projections. For example, the relative errors of focal lengths are around 1.8% under orthogonality error $\sigma = 0.3°$ and pixel noise $\sigma = 1.0$ pixels (Fig. 8 top left), while the focal length errors are around 1.4% when there is same pixel noise $\sigma = 1.0$ pixels but no orthogonality error (Fig. 6 top left).

6.2. Real data

We also applied our method to real images. The first image set consisted of three views of a standing person and a lamp (see Fig. 9). For each pair of images, we applied our algorithm independently, and the results are shown in Table 2. In order to evaluate our results, we used the method in [23] to obtain an over-determined least-squares solution for our camera internal parameters from the noisy measurements. Results from this method are listed in the last column of Table 2 as the ground truth for comparison. The largest relative error in the focal length, in our method, is less than 4.5%. The maximum relative error of principal point is around 5.4%. In addition, the computed polar angle ϕ and the azimuthal angle θ are 44.54° and 33.22°, respectively, while they are 45.09° and 32.97° by using the intrinsic parameters in the last column of Table 2. To verify how these constraints can be used to increase the overall accuracy when more images are used, we applied our method on all the three views and compared with the ground truth as shown in Table 2. As expected, the more images are used, the better is the performance. The errors could be attributed to several sources. Besides noise, nonlinear distortion, and errors in the extracted features, one source is the casual experimental setup using minimal information, which is deliberately targeted to demonstrate a wide spectrum of applications. Despite all these factors, our experimentations indicate that the proposed algorithm provides good results.

6.3. Application to image-based rendering

To demonstrate the strength and the utility of the proposed algorithm, we show two examples for augmented

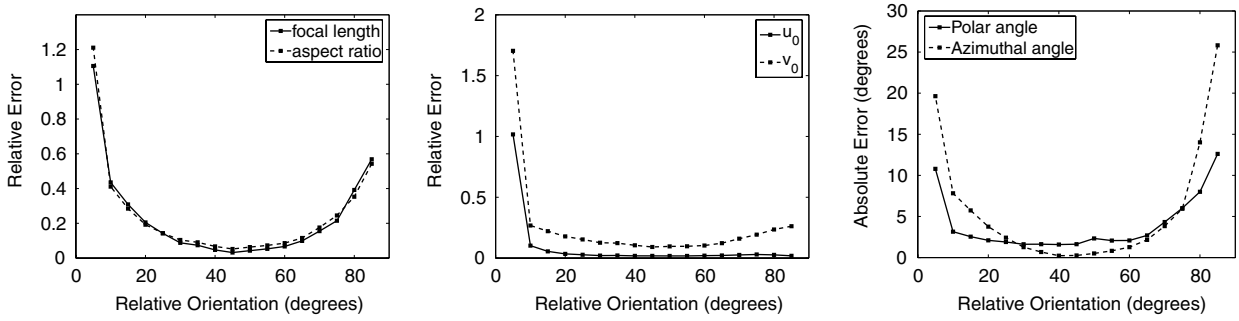

Fig. 7. The performance of our method with respect to the relative orientation between the vertical plane and the image plane.

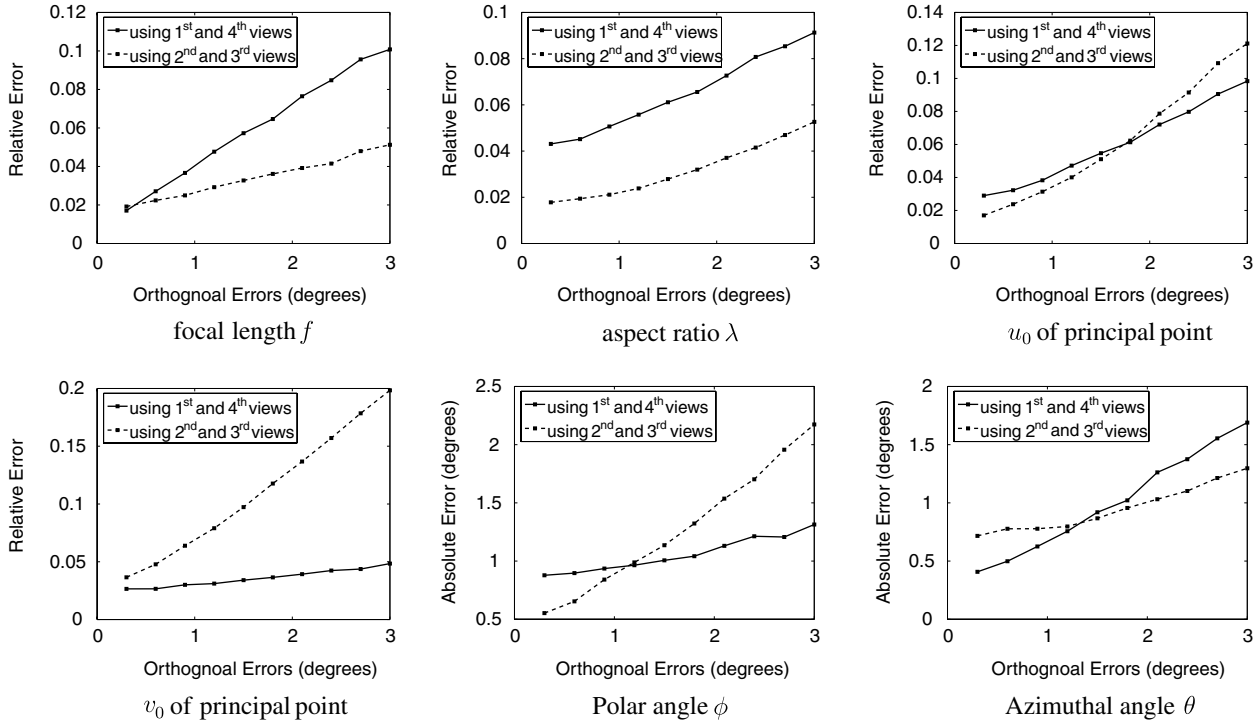

Fig. 8. The performance of our method with respect to the orthogonality errors. The orthogonality errors here refer to errors in estimating the footprints $\mathbf{b}_i$, which enforce the constraint that $\mathbf{t}_i\mathbf{b}_i$ must be orthogonal to the ground plane.

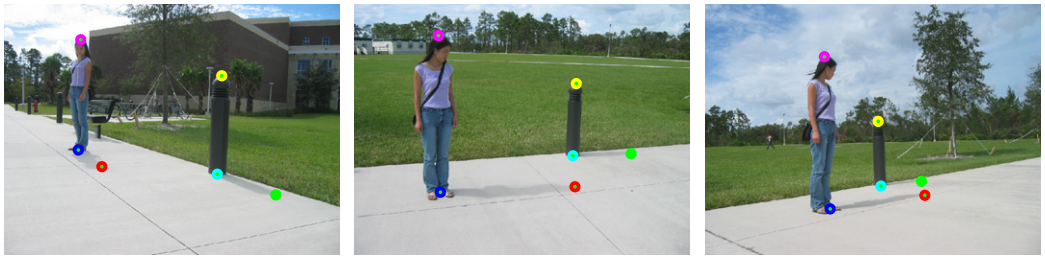

Fig. 9. Three images of a standing person and a lamp. The circle marks in the images are the minimal data computed by the method described in Section 4.

Table 2
Calibration results for the first real image set

Relative error	Test images				[23]
	(1,2)	(1,3)	(2,3)	(1–3)	
f	3200.8 (1.44%)	3137.1 (−0.58%)	3211.1 (1.77%)	3173.2 (0.57%)	3155.3
λf	3206.1 (−3.21%)	3171.0 (−4.27%)	3216.4 (−2.90%)	3267.8 (−1.35%)	3312.5
u_0	1172.4 (0.28%)	1334.9 (5.43%)	1170.1 (0.21%)	1168.2 (0.15%)	1163.6
v_0	895.8 (−0.57%)	902.7 (−0.35%)	901.5 (−0.39%)	906.1 (−0.24%)	913.8

reality by making use of the camera and light source orientation information computed by our method. Given two views as shown in Fig. 10a and b, the computed $\mathbf{K}$ is

$$\mathbf{K} = \begin{bmatrix} 2641.11 & 0 & 991.70 \\ 0 & 2783.97 & 642.28 \\ 0 & 0 & 1 \end{bmatrix},$$

and the computed polar angle ϕ and azimuthal angle θ are 47.99° and 54.91°, respectively. As a result, we can render a virtual teapot with known 3D model in the real scene (b) as shown in (c), where the color characteristics are estimated using the method presented in [5]. Alternatively, we can also follow the method presented in [5] to composite the standing person extracted from (d) with the same person

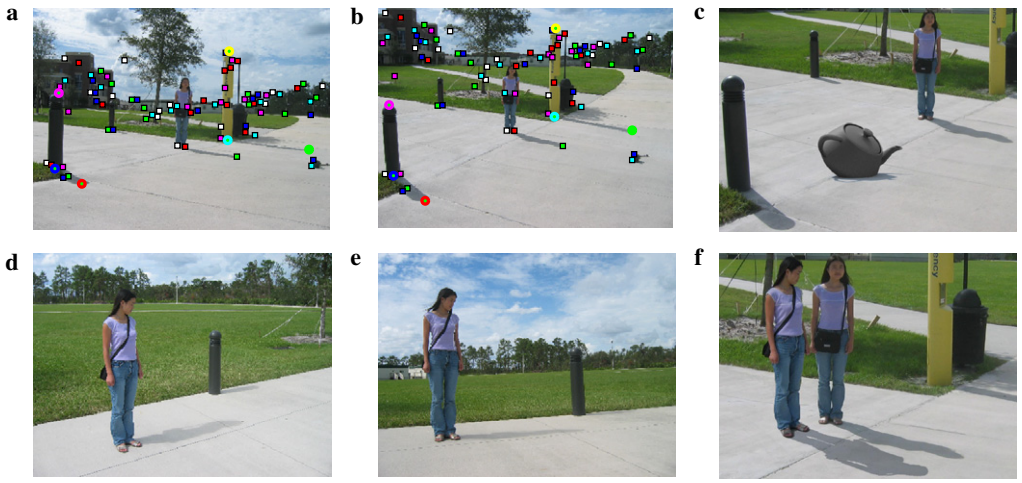

Fig. 10. Image-based rendering Applications. Starting from two views (a) and (b), we first calibrate the camera and compute the light source orientation. The square marks are the corresponding points between the two images, which are computed by using the method in [48]. As a result, we can render a virtual teapot with known 3D model in the real scene (b) shown in (c). Using this computed geometric information, we can also insert a second instance of a person seen in (d) into (b) as shown in (f). (e) The view from light source and is used for shadow map [5].

viewed in (b) and her shadow synthesized from her contour in (e). As a result, we obtain a realistic image in (f) with a person standing next to herself.

7. Conclusion and future work

The proposed calibration technique uses solar shadows which are frequently found in natural environments. The fact that prior knowledge of the 3D coordinates of the 3D stationary points or light source orientation is not required, makes the proposed method a versatile tool that can be used without requiring a precisely machined calibration rig (e.g. grids). It also makes calibration possible when the object is not accessible for measurements, in other words, when the images are taken by other people. More importantly, it relaxes some of the constraints found in other calibration algorithms using vanishing points of orthogonal directions, and hence is especially useful in cases where only limited information is available. This is achieved by enforcing the new inter-image constraints on the image of the absolute conic derived from the *weak pole–polar* relationship. Experimental results show that the method provides very promising solutions even with minimum requirements of two images.

Acknowledgment

The support of the National Science Foundation through the Grant #IIS-0644280 is gratefully acknowledged.

References

[1] M. Agrawal, L. Davis, Complete camera calibration using spheres: Dual space approach, in: Proc. IEEE ICCV, 2003, pp. 782–789.
[2] M. Antone, M. Bosse, Calibration of outdoor cameras from cast shadows, in: Proc. IEEE Int. Conf. Systems, Man and Cybernetics, 2004, pp. 3040–3045.
[3] J. Bouguet, P. Perona. 3d photography on your desk, in: Proc. IEEE ICCV, IEEE Computer Society, 1998, pp. 43–50.
[4] X. Cao, M. Shah. Camera calibration and light source estimation from images with shadows, in: Proc. IEEE CVPR, 2005, pp. 918–923.
[5] X. Cao, Y. Shen, M. Shah, H. Foroosh, Single view compositing with shadows, The Visual Comput. 21 (8) (2005) 639–648.
[6] B. Caprile, V. Torre, Using vanishing points for camera calibration, Int. J. Comput. Vision 4 (2) (1990) 127–140.
[7] Q. Chen, H. Wu, T. Wada, Camera calibration with two arbitrary coplanar circles, in: Proc. ECCV, 2004, pp. 521–532.
[8] C. Colombo, A.D. Bimbo, F. Pernici, Metric 3D reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view, IEEE Trans. Pattern Anal. Mach. Intell. 27 (1) (2005) 99–114.
[9] A. Criminisi, I. Reid, A. Zisserman, Single view metrology, Int. J. Comput. Vision 40 (2) (2000) 123–148.
[10] O. Faugeras, Q.T. Luong, The Geometry of Multiple Images, MIT Press, Cambridge, 2001.
[11] P. Gurdjos, A. Crouzil, R. Payrissat, Another way of looking at plane-based calibration: the centre circle constraint, in: Proc. ECCV, 2002, pp. 252–266.
[12] P. Gurdjos, P. Sturm, Methods and geometry for plane-based self-calibration, in: Proc. IEEE CVPR, 2003, pp. 491–496.
[13] R.I. Hartley, P. Sturm, Triangulation, in: Proc. CAIP, 1995, pp. 190–197.
[14] R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2004.
[15] J. Heikkila, Geometric camera calibration using circular control points, IEEE Trans. Pattern Anal. Mach Intell. 22 (10) (2000) 1066–1077.
[16] T. Horprasert, D. Harwood, L.S. Davis, A statistical approach for realtime robust background subtraction and shadow detection, in: Proc. IEEE ICCV, Frame Rate Workshop, 1999, pp. 1–19.
[17] O. Javed, M. Shah, Tracking and object classification for automated surveillance, in: Proc. ECCV, 2002, pp. 343–357, 2002.
[18] G. Jiang, H.T. Tsui, L. Quan, A. Zisserman, Single axis geometry by fitting conies, IEEE Trans. Pattern Anal. Mach Intell. 25 (10) (2003) 1343–1348.

[19] S. Kang, R.S. Weiss, Can we calibrate a camera using an image of a flat, textureless lambertian surface? in: Proc. ECCV, 2000, pp. 640–653.
[20] A. Klaus, J. Bauer, K. Karner, P. Elbischger, R. Perko, H. Bischof, Camera calibration from a single night sky image, in: Proc. IEEE CVPR, 2004, pp. 151–157.
[21] N. Krahnstoever, P.R.S. Mendonca, Bayesian autocalibration for surveillance, in: Proc. IEEE ICCV, 2005, pp. 1858–1865.
[22] D. Kriegman, P. Belhumeur, What shadows reveal about object structure, in: Proc. ECCV, 1998, pp. 399–414.
[23] D. Liebowitz, A. Zisserman, Combining scene and auto-calibration constraints, in: Proc. IEEE ICCV, 1999, pp. 293–300.
[24] C. Lin, A. Huertas, R. Nevatia, Detection of buildings using perceptual groupings and shadows, 1994, 62–69.
[25] F. Lv, T. Zhao, R. Nevatia, Self-calibration of a camera from video of a walking human, in: Proc. ICPR, 2002, pp. 562–567.
[26] E. Malis, R. Cipolla, Camera self-calibration from unknown planar structures enforcing the multi-view constraints between collineations, IEEE Trans. Pattern Anal. Mach. Intell. 24 (9) (2002) 1268–1272.
[27] P.R.S. Mendonça, Multiview Geometry: Profiles and Self-Calibration. Ph.D. thesis, University of Cambridge, Cambridge, UK, May 2001.
[28] X. Meng, Z. Hu, A new easy camera calibration technique based on circular points, Pattern Recogn. 36 (5) (2003) 1155–1164.
[29] I. Mikic, P. Cosman, G. Kogut, M. Trivedi, Moving shadow and object detection in traffic scenes, in: Proc. ICPR, vol. 1, 2000, pp. 321–324.
[30] S. Nadimi, B. Bhanu, Physical models for moving shadow and object detection in video, IEEE Trans. Pattern Anal. Mach. Intell. 26 (8) (2004) 1079–1087.
[31] T. Okabe, I. Sato, Y. Sato, Spherical harmonics vs. haarwavelets: basis for recovering illumination from cast shadows, in: IEEE Conference on CVPR, 2004, pp. 50–57.
[32] L. Petrovic, B. Fujito, L. Williams, A. Finkelstein, Shadows for cel animation, in: Proc. ACM SIGGRAPH, 2000, pp. 511–516.
[33] M. Pollefeys, R. Koch, L.V. Gool, Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters, Int. J. Comput. Vision 32 (1) (1999) 7–25.
[34] A. Prati, I. Mikic, M. Trivedi, R. Cucchiara, Detecting moving shadows: algorithms and evaluation, IEEE Trans. Pattern Anal. Mach. Intell. 25 (7) (2003) 918–923.
[35] W. Press, B. Flannery, S. Teukolsky, W. Vetterling, Numerical Recipes in C, Cambridge University Press, Cambridge, 1988.
[36] I. Reid, A. North,3D trajectories from a single viewpoint using shadows, in: Proc. of BMVC, 1998.
[37] I. Sato, Y. Sato, K. Ikeuchi, Illumination from shadows, IEEE Trans. Pattern Anal. Mach Intell. 25 (3) (2003) 290–300.
[38] S. Sinha, M. Pollefeys, L. McMillan, Camera network calibration from dynamic silhouettes, in: Proc. IEEE CVPR, 2005, pp. 195–202.
[39] J. Stauder, R. Mech, J. Ostermann, Detection of moving cast shadows for object segmentation, IEEE Trans. Multimedia 1 (1) (1999) 65–76.
[40] P. Sturm, S. Maybank, On plane-based camera calibration: a general algorithm, singularities, applications, in: Proc. IEEE CVPR, 1999, pp. 432–437.
[41] H. Teramoto and G. Xu. Camera calibration by a single image of balls: from conic to the absolute conic, in: Proc. ACCV, 2002, pp. 499–506.
[42] B. Triggs, Autocalibration from planar scenes, in: Proc. ECCV, 1998, pp. 89–105.
[43] B. Triggs, P. McLauchlan, R.I. Hartley, A. Fitzgibbon, Bundle adjustment—a modern synthesis, in: Vision Algorithms: Theory and Practice, 1999, pp. 298–373.
[44] R.Y. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses, IEEE J. Robotics Autom. 3 (4) (1987) 323–344.
[45] L. Van Gool, M. Proesmans, A. Zisserman, Planar homologies as a basis for grouping and recognition, Image Vision Comput. 16 (1998) 21–26.
[46] K.-Y. Wong, R.S. P Mendonca, R. Cipolla, Camera calibration from surfaces of revolution, IEEE Trans. Pattern Anal. Mach Intell. 25 (2) (2003) 147–161.
[47] Y. Wu, H. Zhu, Z. Hu, F. Wu, Camera calibration from the quasi-affine invariance of two parallel circles, in: Proc. ECCV, 2004, pp. 190–202.
[48] J. Xiao, M. Shah, Two-frame wide baseline matching, in: Proc. IEEE ICCV, 2003, pp. 603–609.
[49] H. Zhang, G. Zhang, K.-Y.K. Wong, Camera calibration with spheres: linear approaches, in: Proc. IEEE ICIP, vol. II, 2005, pp. 1150–1153.
[50] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. Pattern Anal. Mach Intell. 22 (11) (2000) 1330–1334.
[51] Z. Zhang, Camera calibration with one-dimensional objects, IEEE Trans. Pattern Anal. Mach Intell. 26 (7) (2004) 892–899.