Camera Calibration and Light Source Estimation from Images with Shadows

Xiaochun Cao and Mubarak Shah
Computer Vision Lab, University of Central Florida, Orlando, FL, 32816

Abstract

In this paper, we describe how camera parameters and light source orientation can be recovered from two perspective views of a scene given only two vertical lines and their cast shadows. Compared to the traditional calibration methods that involve images of some precisely machined calibration pattern, our method uses new calibration objects: the vertical objects and their parallel shadow lines, which are common in natural environments. In addition to the benefit of increasing accessibility of the calibration objects, the proposed method is also especially useful in cases where only limited information is available. To demonstrate the accuracy and the applications of the proposed algorithm, we present results on both synthetic and real images.

1 Introduction

There has been much work on camera calibration, both in photogrammetry and computer vision. Traditional methods (e.g. [6, 10, 13]) typically use a special calibration object with a fixed 3D geometry, and give very accurate results. In some applications, however, it might not be possible to extract camera information off-line by using calibration objects due to the inaccessibility of the camera. Although the recent auto-calibration techniques [7] that aim to compute a metric reconstruction from multiple uncalibrated images avoid the onerous task of calibrating cameras using special calibration objects, they mostly require more than three views and also involve the solution of non-linear problems.

The proposed technique, which uses only two views of a scene containing two vertical objects and their cast shadows, is based on exploiting the priors of a normal camera such that the skew is close to zero and aspect ratio is almost unity as argued in [7]. Instead of assuming them as known in works such as [8, 1], however, we show it is possible to determine them without further assumptions by minimizing the symmetric transfer errors and epipolar distances. Before that, we describe how to express the planar homographies and the fundamental matrix as functions of two components of the Image of Absolute Conic. This proposed method is, therefore, especially useful for cases where only limited information is available.

Another more important advantage of the proposed method is its simplicity and the wide accessibility of the the calibration objects - some vertical objects (e.g. walls, standing people, desks, street lamps, etc.) and their parallel cast shadows illuminated by infinite light source (e.g. sunlit). We admit some recent efforts using architectural buildings [8], surfaces of revolution [11, 4] and circles [3] are toward the similar goal. However, we believe that the alternative vertical objects and their cast shadows are more common in the real world, especially in out-door environments.

Considering also that the appearance of an object greatly depends not only upon the pose of the object but also upon the illumination conditions, the recovery of light source information, similar to the camera calibration, is also crucial in computer vision as well as in computer graphics, especially due to the recent interest in Image-Based Rendering (IBR) techniques. In this work, therefore, we focus on a typical outdoor scene with several vertical objects lit by a distant sunlight, although this proposed method is not that restricted. For example, our method also works for the case with two vertical objects and a finite vanishing point along a direction orthogonal to the vertical one. We show two views of such scenes are enough to calibrate the camera and recover the orientation of the light source. Since the developed technique requires no knowledge of the 3D coordinates of the feature points of the vertical objects, it is well-suited for IBR applications. Two examples will be used to show how to make use of the camera and light source information, and demonstrate the strength and applicability of this methodology.

2 Preliminaries

2.1 Pin-hole Camera Model

A pin-hole camera, based on the principle of collinearity, projects a region of $\mathbb{R}^3$ lying in front of the camera into a region of the image plane $\mathbb{R}^2$. As is well known, a 3D point $M = [X\ Y\ Z\ 1]^T$ and its corresponding projection

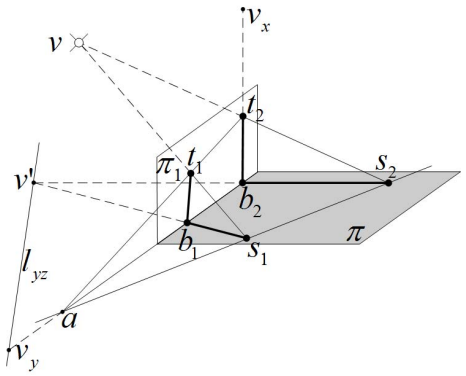


Figure 1. Basic geometry of a scene with two vertical lines t_1b_1 and t_2b_2 casting shadows s_1b_1 and s_2b_2 on the ground plane π by the distant light source v.

$m = [u\ v\ 1]^T$ in the image plane is related via a 3×4 matrix $\mathbf{P}$ as

$$m \sim \underbrace{K[r_1\ r_2\ r_3\ t]}_{\mathbf{P}} M, \quad K = \begin{bmatrix} f & \gamma & u_0 \\ 0 & \lambda f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (1)$$

where $\sim$ indicates equality up to multiplication by a non-zero scale factor, the r_1, r_2, r_3 are the columns of the 3×3 rotation matrix R, $t = -RC$, with $C = [C_x\ C_y\ C_z]^T$ being the relative translation between the world origin and camera center, is the translation vector, and K is the 3×3 camera intrinsic matrix including five parameters: focal length f, skew γ, aspect ratio λ and principal point at $(u_0,\ v_0)$.

2.2 Scene Configuration

We first examine the scenes containing two vertical lines and their cast shadows on the ground plane. The basic geometry is shown in Fig. 1. Note that this figure shows the projections of the world points in the image planes denoted by corresponding lower-case characters. For example, the world point B_2 (not shown in Fig. 1) is mapped to b_2 in the image plane. Without loss of generality, we choose the world coordinate frame as follows: origin at B_2, X-axis along the line B_2T_2 with the positive direction towards T_2, Y-axis along the line B_1B_2 with the negative direction towards B_1, and the Z-axis given by the right-hand rule.

2.3 Constraints from A Single View

In the following, we explore the constraints available from a single view, given the above configuration. Based on the world coordinate frame described above, we can compute the vanishing point v_x along the x-axis (i.e. vertical) direction by intersecting the two vertical lines t_1b_1 and t_2b_2. Since the light source, v, is at infinity or distant, the two shadow lines S_2B_2 and S_1B_1 must be parallel in the 3D world. In other words, the two imaged parallel shadow lines, s_1b_1 and s_2b_2, will intersect in the image space at the vanishing point v'.

From the pole-polar relationship with respect to the *Image of the Absolute Conic* ω – an imaginary point conic directly related to the camera internal matrix K in (1) as $\omega = K^{-T}K^{-1}$[7]: the vanishing point v_x of the normal direction to a plane (ground plane π in our case) is the pole to the polar which is the vanishing line l_{yz} of the plane,

$$v_y \times v' = l_{yz} = \omega v_x, \qquad (2)$$

where v_y is the vanishing point along the y-axis. Equation (2) can be rewritten, equivalently, as two constraints on ω:

$$v'^T \omega v_x = 0, \qquad (3)$$

$$v_y^T \omega v_x = 0. \qquad (4)$$

In our case, we only have the constraint (3) since we can not determine v_y yet. Without further assumptions, we are unlikely to extract more constraints on K from a single view of such a scene shown in Fig. 1.

Before we move to our method, we do want to mention some possible configurations that may provide more constraints, although we will not make use of such constraints. One possibility is to assume that the two vertical lines t_1b_1 and t_2b_2 have the same lengths, in which case v_y can be directly computed as $v_y = (t_1 \times t_2) \times (b_1 \times b_2)$. Other possibilities include utilizing the knowledge of the orientation of the light source v, or making use of the ratios of lengthes such as t_1b_1/t_2b_2 and t_1b_1/b_1b_2. However, too many assumptions limit the applicabilities in the real world.

3 Our Method

Our proposed method aims to solve the relatively more general problem using two views without any further assumptions. The basic idea of our method is to define two camera matrices P and P' corresponding to the two views as functions of ω_{12} and ω_{22}, two elements of the ω. The reason we choose ω_{12} and ω_{22} as variables will be explained in section 3.2. As a result, we can compute ω_{12} and ω_{22} by minimizing the symmetric transfer errors of the geometric distances and the epipolar distances. Therefore, both camera intrinsic and external parameters can be recovered since P and P' depend only on ω_{12} and ω_{22}.

3.1 Extra Constraints from the Second View

The second view can be easily used to get the second constraint from equation (3). Beyond that, we explore

here the third constraint based on the invariance property of cross-ratio under projective transformation.

Geometrically, equation (4) can be interpreted as v_y lies on the line ωv_x. Considering also that v_y lies on the imaged y-axis $b_1 b_2$, we can express v_y as a function of ω:

$$v_y = [b_1 \times b_2]_\times \omega v_x, \tag{5}$$

where $[\cdot]_\times$ is the notation for the skew symmetric matrix characterizing the cross product. As shown in Fig. 1, the four points a, b_1, b_2, and v_y, are collinear, and their cross-ratio is preserved under the perspective projection. Thus we have the following equality between two given images

$$\{v_y, b_2; b_1, a\}^1 = \{v_y, b_2; b_1, a\}^2, \tag{6}$$

where $\{\cdot,\cdot;\cdot,\cdot\}^i$ denotes the cross ratio of four points, and the superscripts indicate the images in which the cross ratios are taken. This gives us the third constraint on $\omega = K^{-T}K^{-1}$, which can be expanded up to a scale as:

$$\omega \sim \begin{bmatrix} 1 & -\frac{\gamma}{f\lambda} & \frac{\gamma v_0 - \lambda f u_0}{f\lambda} \\ * & \frac{f^2+\gamma^2}{f^2\lambda^2} & -\frac{\gamma^2 v_0 - \gamma\lambda f u_0 + v_0 f^2}{f^2\lambda^2} \\ * & * & \frac{v_0^2(f^2+\gamma^2) - 2\gamma v_0 \lambda f u_0}{f^2\lambda^2} + f^2 + u_0^2 \end{bmatrix}, \tag{7}$$

where the lower triangular elements are denoted by $*$ to save the space since ω is symmetric. Therefore, we can define ω by a 6D vector with five unknowns as:

$$w \sim [1, \omega_{12}, \omega_{22}, \omega_{13}, \omega_{23}, \omega_{33}]^T, \tag{8}$$

where ω_{ij} denotes the element in i^{th} row and j^{th} column of ω in (7). If we assume a simplified camera model with zero skew and unit aspect ratio, theoretically, these three known constraints are sufficient to solve for the three unknowns: focal length f, principal point coordinates u_0 and v_0.

3.2 Defining P & P' as functions of ω_{12} & ω_{22}

In practice, however, it may be more interesting to fully calibrate the camera. Instead of treating some internal parameters (for example γ and λ) as known or constant, we first define both the camera internal parameters and external parameters as functions of ω_{12} and ω_{22}. The reason we choose ω_{12} and ω_{22} as variables is that they embrace the experimental knowledge of a camera model. In other words, ω_{12} is scaled γ by $1/f$ and thus very close to zero, while ω_{22} is close to $1/\lambda^2 \approx 1$. As a result, we can compute ω_{12} and ω_{22} by enforcing inter-image planar homography and epipolar geometric constraints as explained later.

Since there are three constraints (two from equation (3) and one from equation (6)) on ω, we can compute ω_{13}, ω_{23} and ω_{33} as functions of ω_{12} and ω_{22}. Without difficulty, we then uniquely extract the intrinsic parameters from ω,

$$\lambda = \sqrt{1/(\omega_{22} - \omega_{12}^2)}, \tag{9}$$

$$v_0 = (\omega_{12}\omega_{13} - \omega_{23})/(\omega_{22} - \omega_{12}^2), \tag{10}$$

$$u_0 = -(v_0\omega_{12} + \omega_{13}), \tag{11}$$

$$f = \sqrt{\omega_{33} - \omega_{13}^2 - v_0(\omega_{12}\omega_{13} - \omega_{23})}, \tag{12}$$

$$\gamma = -f\lambda\omega_{12}. \tag{13}$$

After expressing the camera internal parameters, we can compute camera external parameters as follows. As is known [6, 5], the first column p_1 of the projection matrix P in equation (1) is the scaled v_x, and the the second column $p_2 \sim v_y$. For example, v_x is the projection in the image plane of the infinite 3D point $X_\infty = [1\ 0\ 0\ 0]^T$,

$$v_x \sim [p_1\ \ p_2\ \ p_3\ \ p_4][1\ \ 0\ \ 0\ \ 0]^T = p_1. \tag{14}$$

By expanding equation (1), We have

$$\begin{aligned} p_1 &= \begin{bmatrix} K_1^T r_1 & \lambda f r_{21} + v_0 r_{31} & r_{31} \end{bmatrix}^T, \\ p_2 &= \begin{bmatrix} K_1^T r_2 & \lambda f r_{22} + v_0 r_{32} & r_{32} \end{bmatrix}^T, \end{aligned} \tag{15}$$

where K_1 is the first row of camera internal matrix K, and r_k are the columns and r_{ij} are the components of the rotation matrix $R = R_z R_y R_x$.

After simple algebraic derivations, three rotation angles can be expressed as functions of camera intrinsic parameters as

$$\theta_z = \tan^{-1} \frac{f(v_{xy} - v_0)}{\lambda f(v_{xx} - u_0) - \gamma(v_{xy} - v_0)}, \tag{16}$$

$$\theta_y = \tan^{-1} \frac{\lambda f \sin(\theta_z)}{v_0 - v_{xy}}, \tag{17}$$

$$\theta_x = \tan^{-1} \frac{\lambda f \cos(\theta_z)/\cos(\theta_y)}{v_{yy} - v_0 - \lambda f \tan(\theta_y)\sin(\theta_z)}, \tag{18}$$

where $(v_{xx}\, v_{xy})$ are the coordinates of v_x, and $(v_{yx}\, v_{yy})$ are the coordinates of v_y. Similar to the work in [1], translation vectors t can also be computed up to a scale.

3.3 Solving for Camera Calibration

Now we have expressed all the camera parameters, and hence camera matrices P and P', as functions of ω_{12} and ω_{22}. Therefore, we can compute ω_{12} and ω_{22} by enforcing both the strong and weak inter-image constraints that minimize the symmetric transfer errors of geometric distances.

The strong constraints are typically planar homographies that have a one to one mapping, and the weak one is often the epipolar constraint. Obviously, we have two dominant planes π and π_1 as shown in Fig. 1. The two inter-frame

planar homographies H_π and H_{π_1} corresponding to π and π_1 can be computed as

$$H_\pi = [p'_2 \;\; p'_3 \;\; p'_4][p_2 \;\; p_3 \;\; p_4]^{-1}, \quad (19)$$

$$H_{\pi_1} = [p'_1 \;\; p'_2 \;\; p'_4][p_1 \;\; p_2 \;\; p_4]^{-1}, \quad (20)$$

where p'_i and p_i denote the i^{th} columns of P' and P respectively. For the corresponding points that do not lie on either plane π or π_1, we enforce the weak epipolar constraints on them. The epipolar constraint is encapsulated in the algebraic representation by the fundamental matrix F that can be computed as

$$F = [e']_\times P' P^+, \quad (21)$$

where P^+ is the pseudo-inverse of P, i.e. $PP^+ = \mathbf{I}$, and the epipole $e' = P'C$ where C is the null-vector of P', namely the camera center, defined by $PC = 0$.

Finally, we can determine the two variables ω_{12} and ω_{22} by enforcing the above constraints, i.e. minimizing the following symmetric transfer errors of the geometric distances and epipolar distances

$$\lambda_1 \sum_{i=1}^{N_\pi} (d_1(x_i, H_\pi^{-1} x'_i)^2 + d_1(x'_i, H_\pi x_i)^2) +$$
$$\lambda_2 \sum_{j=1}^{N_{\pi_1}} (d_1(x_j, H_{\pi_1}^{-1} x'_j))^2 + d_1(x'_j, H_{\pi_1} x_j)^2) +$$
$$\lambda_3 \sum_{k=1}^{N_g} (d_2(x'_k, F x_k))^2 + d_2(x_k, F^{-1} x'_k)^2). \quad (22)$$

where $d_1(\cdot,\cdot)$ is the Euclidean distance between two points, $d_2(\cdot,\cdot)$ is the Euclidean distance from a point to a line, λ_i are the weights, and N_* are the numbers of matching points coincide with different constraints. The initial estimates for ω_{12} and ω_{12} are zero and one respectively.

3.4 Light Source Orientation Estimation

After calibrating these cameras, we have no difficulty in estimating the light source position and orientation by using the triangulation method [7]. Since in our case the light source is far away, however, we only need to measure azimuthal angle θ in the YZ plane with the Y-axis and the polar angle ϕ with the X-axis as shown in Fig. 1.

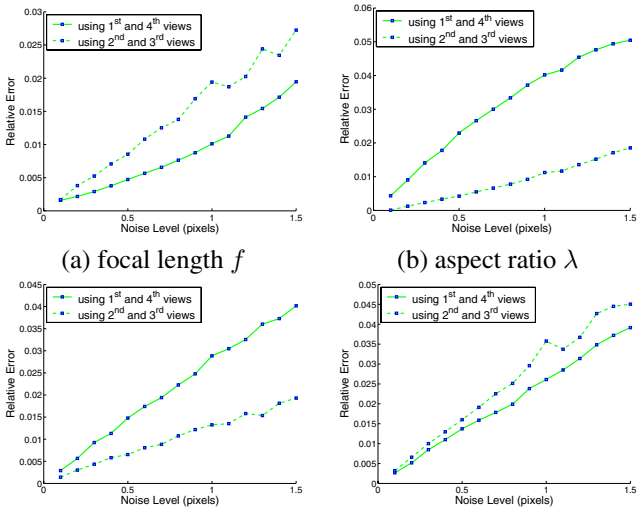
$$\phi = \cos^{-1} \frac{v_x^T \omega v}{\sqrt{v^T \omega v}\sqrt{v_x^T \omega v_x}}, \theta = \cos^{-1} \frac{v_y^T \omega v'}{\sqrt{v'^T \omega v'}\sqrt{v_y^T \omega v_y}}. \quad (23)$$

4 Experimental Results

The proposed method aims to directly calibrate cameras for applications where it is difficult to calibrate cameras beforehand using special calibration pattern with known geometry and where not enough numbers of views are available to employ self-calibration methods. In our experiments, therefore, we focus on the cases with minimal information. The minimal information is nothing but six points in two views as described in section 2.2.

View	camera position	"at" position
1^{st}	(40 -10 100+randnom(1))	(0 100 0)
2^{nd}	(40 150 100+randnom(1))	(0 100 0)
3^{rd}	(100 -10 100+randnom(1))	(0 100 0)
4^{th}	(100 150 100+randnom(1))	(0 100 0)

Table 1. Parameters for four viewpoints.



(a) focal length f (b) aspect ratio λ

(c) u_0 of principal point (d) v_0 of principal point

Figure 2. Performance vs. noise (in pixels) averaged over 1000 independent trials.

4.1 Computer Simulation

The simulated camera has a focal length of $f = 1000$, aspect ratio of $\lambda = 1.06$, skew of $\gamma = 0.06$, and the principal point at $u_0 = 8$ and $v_0 = 6$. The two vertical objects have lengths 100 and 80 pixels respectively, and the distance between the two vertical objects is 75 pixels. The polar angle $\phi = \arctan 0.5$ and the azimuthal angle $\theta = 60°$. In the experiments presented herein, we generated four views with camera and "at" positions listed in Table 1. Note that we follow camera coordinate specification in OpenGL fashion. Therefore, $at - camera$ is the principal view direction.

We used the two combinations of image pairs (views) in Table 1. The first combination composes of 1^{st} and 4^{th}

Figure 3. Three images of a standing person and a lamp. The circle marks in the images are the minimal data. The square marks are the corresponding points, computed by using method [12], between the last two images, which are used to compute the epipolar distances in equation (22).

Error	Image Pair			
	(1,2)	(1,3)	(2,3)	[8]
f	3203.1	3179.8	3208.0	3155.3
relative error	1.51%	0.78%	1.67%	
λf	3208.3	3185.0	3213.2	3312.5
relative error	-3.15%	-3.85%	-2.99%	
u_0	1176.3	1299.0	1170.5	1163.6
relative error	0.40%	4.29%	0.22%	
v_0	896.1	900.9	902.2	913.8
relative error	-0.56%	-0.41%	-0.37%	
γ	-0.67	-0.66	-0.55	-0.62

Table 2. Results for the first real image set.

views, while the second one includes 2^{nd} and 3^{rd} views. Gaussian noise with zero mean and a standard deviation of $\sigma \leq 1.5$ was added to the projected image points. The estimated camera parameters were then compared with the ground truth. As argued by [9, 14], the relative difference with respect to the focal length rather than the absolute error is a geometrically meaningful error measure. Therefore, we measured the relative error of f, u_0 and v_0 with respect to the f while varying the noise level from 0.1 pixels to 1.5 pixels. For each noise level, we perform 1000 independent trials, and the results shown in Fig. 2 are the average. For the aspect ratio λ, we measure the relative error w.r.t. itself. As pointed out in [7], γ will be zero for most normal cameras and can take non-zero values only in certain unusual instances (i.e. taking an image of an image). Without surprise, we found in our experiment that the experimental results are very insensitive to the variable ω_{12} since $\omega_{12} = -\gamma/(f\lambda) \approx 0$ in most cases and equals to $-5.6604e-5$ in our case. In other words, small amount of noise will overcome the valid information to extract the skew parameter γ and the result of γ is not very meaningful. Results of other four camera internal parameters are shown in Fig. 2. Errors increase almost linearly with respect to the noise level. When we add more noise, the relative errors of focal lengths keep increasing until it reaches 1.95% for the first combination and 2.72% for the second one when $\sigma = 1.5$. The maximum relative error of aspect ratio is 5.51% for the first combination, 1.47% for the second combination. The maximum relative errors of principal points are around 4.02% for u_0 and about 4.51% for v_0.

4.2 Real Data

We also applied our method on real images. The first image set consisted of three views of a standing person and a lamp, which provided two vertical lines for camera calibration (see Fig. 3). 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 obtain a least-squares (non-natural camera) solution for internal parameters from over-determined noisy measurements, i.e. five images with three mutually orthogonal vanishing points per view, using the constraints described in [8]. We compared our results to those listed in the last column in Table 2. The largest relative error of the focal length, in our case, is less than 4%. The maximum relative error of principal point is around 4.3%. In addition, the computed polar angle ϕ and azimuthal angle θ are 44.45 and 33.17 degrees respectively, while they are 45.09 and 32.97 degrees by using the camera intrinsic parameters in the last column of Table 2. The errors could be attributed to several sources. Besides noise, non-linear distortion and imprecision of the extracted features, one source is the casual experimental setup using minimal information, which is deliberately targeted for a wide spectrum of applications. Despite all these factors, our experimentations indicate that the proposed algorithm provides good results.

Application to image-based rendering: To demonstrate the strength and applicability of the proposed algorithm, we show two examples for augmented reality by making use of the camera and light source orientation information computed by our method. Given two views as shown in Fig. 4 (a) and (b), the computed camera internal matrix is

$$K = \begin{bmatrix} 2641.08 & 0.03 & 991.85 \\ 0 & 2783.95 & 642.30 \\ 0 & 0 & 1 \end{bmatrix},$$

and the computed polar angle ϕ and azimuthal angle θ are 48.32 and 54.91 degrees respectively. As a result, we can render a virtual teapot with known 3D model into the real scene (b) shown in (c), the color characteristic is estimated using methods presented in [2]. Alternatively, we can also follow the method presented in [2] to composite the standing person extracted from (d) into (b) and synthesize its shadow using the contour of the person in (e). Note that

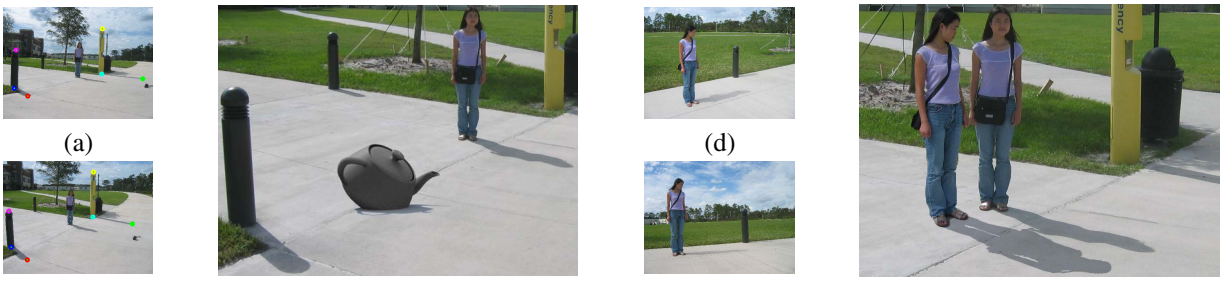


Figure 4. Image-based rendering Applications. Starting from two views (a) and (b), we first calibrate the camera and compute the light source orientation. Then, we can render a virtual teapot with known 3D model into (b) shown in (c). Utilizing this computed geometric information, we can also insert another person (d) into (b) as shown in (f).

(e) is the image taken along the lighting direction, not necessarily the view from the light source.

5 Conclusion and Future Work

The proposed calibration technique uses images of vertical objects and their parallel cast shadows which are frequently found in natural environment. The fact that prior knowledge of the 3D coordinates of the vertical objects is not required, makes the method a versatile tool that can be used without requiring a precisely machined calibration rig (e.g. grids), and also makes calibration possible when the object is not accessible for measurements, in other words, when the images are taken by other people. Moreover, our method alleviates the limitation of a simplified camera model for cases where only limited information is available. This is achieved by enforcing inter-image homography and epipolar geometric constraints, and exploiting the property of a normal camera that the skew is close to zero and aspect ratio is almost unity. Experimental results show that the method provides very promising solutions even with minimum requirements of two images and six correspondences.

Acknowledgment

This material is based upon work funded in part by the U.S. Government. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Government.

References

[1] X. Cao and H. Foroosh. Simple calibration without metric information using an isosceles trapezoid. In *Proc. ICPR*, pages 104–107, 2004.

[2] X. Cao and M. Shah. Creating realistic shadows of composited objects. In *Proc. Wkshp. App. of Comp. Vis.*, pages 294–299, 2005.

[3] Q. Chen, H. Wu, and T. Wada. Camera calibration with two arbitrary coplanar circles. In *Proc. ECCV*, pages 521–532, 2004.

[4] C. Colombo, A. Bimbo, and 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):99–114, 2005.

[5] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *Int. J. Comput. Vision*, 40(2):123–148, 2000.

[6] O. Faugeras. *Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[8] D. Liebowitz and A. Zisserman. Combining scene and autocalibration constraints. In *Proc. IEEE ICCV*, pages 293–300, 1999.

[9] B. Triggs. Autocalibration from planar scenes. In *Proc. ECCV*, pages 89–105, 1998.

[10] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. of Robotics and Automation*, 3(4):323–344, 1987.

[11] K.-Y. Wong, R. Mendonca, and R. Cipolla. Camera calibration from surfaces of revolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):147–161, 2003.

[12] J. Xiao and M. Shah. Two-frame wide baseline matching. In *Proc. IEEE ICCV*, pages 603– 609, 2003.

[13] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, 2000.

[14] Z. Zhang. Camera calibration with one-dimensional objects. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(7):892–899, 2004.