PERGAMON

Pattern Recognition 35 (2002) 2937–2948

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Self-calibration with varying focal length from two images obtained by a stereo head

J.-S. Liu[a,*], J.-H. Chuang[b]

[a] *Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung, Taiwan, ROC*
[b] *Department of Computer and Information Science, National Chiao Tung University, Hsinchu 30010, Taiwan, ROC*

Received 13 October 2000; accepted 30 October 2001

Abstract

The self-calibration approach based on the absolute conic or its dual, the absolute dual quadric, has the merit of allowing the intrinsic camera parameters to vary in image sequence. In this paper, we show that certain linear equations resulting from the infinity homography can be added to a system of originally undetermined linear equations to find the absolute dual quadric for a stereo head. The special stereo configurations considered here allow one camera to rotate around either one or two of the coordinate axes defined by another camera. Experiments with synthetic images show that satisfactory results can be obtained with only the proposed linear methods. For real images obtained with non-ideal stereo configurations, and possibly with measurement noises, results obtained from the proposed linear methods may serve as reasonable initial guesses for some non-linear optimization procedure. © 2002 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Self-calibration; Stereo vision; Absolute conic; Infinity homography

1. Introduction

Self-calibration of a camera from images has been an important research topic on computer vision over the last few years since it may reduce the need of off-line calibration and increases on-line flexibility. It is shown in Refs. [1,2] that general projective reconstructions, i.e., the simplest type of self-calibration, can be obtained easily using two or more uncalibrated projective images. Recently, more and more researchers pay their attention to possible ways of upgrading these reconstructions from projective to metric. Faugeras et al. [3] proposed a robust self-calibration method using the Kruppa equations to impose constraints on the fixed internal parameters obtained from fundamental matrix. A number of approaches based on similar concepts to self-calibration have been developed [4–7].

Instead of using the Kruppa equations, some widely accepted approaches [8–10] are based on the absolute quadric, a concise parameterization of the absolute conic first introduced by Heyden [8] and called Kruppa constraint, and later proposed by Triggs formally [11]. By means of this parametric representation, it is shown that the self-calibration can be done even if the camera intrinsic parameters are allowed to vary in a motion sequence. On the other hand, based on the infinity homography, approaches of reconstruction from projective to affine and finally to Euclidian space have also been widely adopted in the last few years [12–14]. In fact, some researchers have dealt with the calibration problem using both the absolute quadric and the infinity homography for some special camera motions [15].

It is shown in Ref. [9] that, under the condition that intrinsic camera parameters, except for the focal length, are known, a linear solution of the varying focal length together with the location of a particular affine structure can be obtained. The linear solution can then be used to initialize the corresponding non-linear optimization procedure. However, it is shown in Ref. [14] that even if all intrinsic parameters are constant, in the special case of having only two images obtained with the self-calibration approach, only one modulus constraint for the infinity homography exists that the

* Corresponding author.
E-mail addresses: jsliu@mail.cyut.edu.tw (J.S. Liu), jchuang@cis.nctu.edu.tw (J.H. Chuang).

PII: S0031-3203(01)00236-9

plane at infinity cannot be determined. Subsequently, a scene constraint obtained from vanishing points is added to solve the self-calibration problem for such a constant camera.

In this paper, we show that, not only for a constant camera but also for that allowing its focal length to vary, without additional scene constraint as required in Ref. [14], certain linear equations resulting from the infinity homography can be added to the system of undetermined linear equations in Ref. [9] to provide a closed form solution to the self-calibration problem, given only two images obtained from certain camera motions, or equivalently from special configurations of a stereo head. As part of an optimal approach, results obtained by the proposed linear approach can be improved further by a subsequent nonlinear optimization procedure.

The paper is organized as follows. In Section 2, some background geometry and notation are introduced. Section 3 describes the general self-calibration based on the absolute quadric and the infinity homography. Then in Section 4, the linear solutions based on the infinity homography constraints are introduced for special configurations of a stereo head. Following that, Section 5 is devoted to a brief summary of the associated 3D metric reconstruction. Some experimental results are given in Section 6. Finally, we draw conclusions in Section 7.

2. Background geometry and notation

In this section, a brief review is given for the classical projective geometry notions of infinity homography, plane at infinity, absolute conic, and their relationships to camera calibration.

2.1. Projection matrix and infinity homography

A basic projection procedure of scene points onto an image by a perspective camera can be described as

$$m \propto PM, \tag{1}$$

where $\propto$ denotes the equality up to a scaling factor, P is the 3×4 projection matrix, $M=[X\ \ Y\ \ Z\ \ 1]^{\mathrm{T}}$ and $m=[x\ \ y\ \ 1]^{\mathrm{T}}$ represent the homogeneous coordinates of a 3D world point and an image point, respectively. Due to the stratums of space, the projection process should be represented by means of its corresponding projection matrix in the space under consideration.

For Euclidean space, the projection matrix can be represented as

$$P_{euc} = KP_0T = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0_3^{\mathrm{T}} & 1 \end{bmatrix}, \tag{2}$$

where T represents the transformation of coordinate systems from world to the camera-centered system, P_0 denotes the perspective projection and K is the camera matrix consisting of the intrinsic parameters of camera. In the camera matrix, f_x and f_y are the focal lengths measured in width and in height of the pixels in the image, respectively, s is a factor measuring the skew of the two image axes, and u_0 and v_0 are the image coordinates of the principal point.

For the projective space, on the other hand, the projection matrix can be represented as

$$P_{proj} = [H|e_r], \tag{3}$$

where e_r is the epipole, and H, the Homography, describes the projection from a particular reference plane to the image plane, as discussed next.

Given a reference plane $\Pi = [\pi^{\mathrm{T}}\ \ 1]^{\mathrm{T}} \triangleq [\pi_1\ \ \pi_2\ \ \pi_3\ \ 1]^{\mathrm{T}}$ in the 3D space, a point $M_\Pi = [m_\Pi^{\mathrm{T}}\ 1]^{\mathrm{T}}$ is said to lie on this plane if and only if $\Pi^{\mathrm{T}}M_\Pi = \pi^{\mathrm{T}}m_\Pi + 1 = 0$. Specifically, since $\pi^{\mathrm{T}}m_\Pi = -1$, the relationship can be represented as

$$M_\Pi = \begin{bmatrix} m_\Pi \\ 1 \end{bmatrix} = \begin{bmatrix} m_\Pi \\ -\pi^{\mathrm{T}}m_\Pi \end{bmatrix} = \begin{bmatrix} I_{3\times 3} \\ -\pi^{\mathrm{T}} \end{bmatrix} m_\Pi. \tag{4}$$

Hence, the projection process which maps the 3D point M_Π to its image point $\tilde{m}_\Pi$ by the projective projection matrix can be described as

$$\tilde{m}_\Pi \propto P_{proj}M_\Pi = [H|e_r] \begin{bmatrix} I_{3\times 3} \\ -\pi^{\mathrm{T}} \end{bmatrix} m_\Pi \tag{5}$$

or

$$\tilde{m}_\Pi \propto [H - e_r\pi^{\mathrm{T}}]m_\Pi \tag{6}$$

Thus,

$$H' \triangleq H - e_r\pi^{\mathrm{T}} \tag{7}$$

represents the homography between M_Π and $\tilde{m}_\Pi$. If the plane Π is chosen to be $[0\ \ 0\ \ 0\ \ 1]^{\mathrm{T}}$, the corresponding homography is simply given by H. This is the homography denoted in the projective projection matrix (3). On the other hand, the infinity homography is denoted as another special homography which describes the transformation from the plane at infinity, Π_∞, to the image plane:

$$H^\infty = H - e_r\pi_\infty^{\mathrm{T}} \tag{8}$$

where π_∞ is the vector consisting of the first three elements of $[\pi_{\infty_1}\ \ \pi_{\infty_2}\ \ \pi_{\infty_3}\ \ 1]$ giving the location of Π_∞. Details about Π_∞ are given in the next subsection.

2.2. Plane at infinity and absolute conic

The plane at infinity, or the infinity plane, Π_∞, is the plane expressed as $X_4 = 0$ in an affine frame and is setwise invariant under Euclidean motions i.e., any rigid motion of a camera will not change its relative position and orientation with respect to Π_∞.

The absolute conic, Ω, is a point conic on Π_∞ defined as $X_1^2 + X_2^2 + X_3^2 = 0$ with $X_4 = 0$, containing only imaginary points [16]. As its dual, the absolute dual quadric is denoted as Ω^*. A special property associated with Ω is that if camera parameters do not change, then the image of Ω, ω, and its

dual, ω^*, will also stay the same for all views. In particular, for a Euclidean representation of the world, such a property can be expressed as

$$\omega^* \propto P_{euc}\Omega^*_{euc}P^{\mathrm{T}}_{euc} = K[R^{\mathrm{T}}|t]\begin{bmatrix} I_{3\times 3}\,0_3 \\ 0_3^{\mathrm{T}}\,0 \end{bmatrix}\begin{bmatrix} R \\ t \end{bmatrix}K^{\mathrm{T}} = KK^{\mathrm{T}}. \quad (9)$$

In cases which allow variable intrinsic camera parameters, there is a particularly useful property of ω_i^*, i.e.,

$$\omega_i^* = K_iK_i^{\mathrm{T}} \propto P_i\Omega^*P_i^{\mathrm{T}} \quad (10)$$

is satisfied for all views (i's). According to Eq. (10), constraints on the intrinsic camera parameters associated with K_i can therefore be transformed to those on elements of ω_i^*. This actually provides a basis for the self-calibration.

3. Self-calibration

The absolute dual quadric and the infinity homography are the basis of the self-calibration since images of the former encode camera matrices for all views while the latter encodes camera rotations. By these parameterizations, location of the plane at infinity as well as the intrinsic camera parameters can be obtained. Thus, the projection matrix for 3D reconstructions in metric space can be obtained. In this section, a general approach to solving the self-calibration problem based on the absolute dual quadric is briefly reviewed. Then, in order to overcome the difficulty induced by the special case of using only two images that the solution in general cannot be determined uniquely, constraints of the infinity homography constraints are derived in the next section for special configurations of a stereo head.

Consider the absolute dual quadric given in Eq. (10). Starting from its Euclidean representation, such a quadric can eventually be expressed in the projective space as

$$\begin{aligned}\omega^* &\propto P_{euc_i}\Omega^*_{euc}P^{\mathrm{T}}_{euc_i} \\ &= (P_{proj_i}T_{PM}^{-1})\Omega^*_{euc}(T_{PM}^{-\mathrm{T}}P^{\mathrm{T}}_{proj_i}) \\ &= P_{proj_i}(T_{PM}^{-1}\Omega^*_{euc}T_{PM}^{-\mathrm{T}})P^{\mathrm{T}}_{proj_i} \\ &= P_{proj_i}\Omega^*_{proj}P^{\mathrm{T}}_{proj_i}, \end{aligned} \quad (11)$$

where P_{proj_i} denotes projection matrix in the projective space for the ith image,

$$T_{PM} = \begin{bmatrix} K^{-1} & 0 \\ \pi_\infty^{\mathrm{T}} & 1 \end{bmatrix} \quad (12)$$

is the transformation matrix to upgrade the geometry from projective to metric Ref. [17], and

$$\Omega^*_{proj} = T_{PM}^{-1}\Omega^*_{euc}T_{PM}^{-\mathrm{T}} = \begin{bmatrix} KK^{\mathrm{T}} & -KK^{\mathrm{T}}\pi_\infty \\ -\pi_\infty^{\mathrm{T}}KK^{\mathrm{T}} & \pi_\infty^{\mathrm{T}}KK^{\mathrm{T}}\pi_\infty \end{bmatrix} \quad (13)$$

is the absolute dual quadric in the projective space.

In particular, if the world frame is aligned with the first camera, i.e., $P_{proj_1} = [I|0]$, then we have

$$K_iK_i^{\mathrm{T}} \propto P_{proj_i}\Omega^{*'}_{proj}P^{\mathrm{T}}_{proj_i} \quad (14)$$

with

$$\Omega^{*'}_{proj} = \begin{bmatrix} K_1K_1^{\mathrm{T}} & -K_1K_1^{\mathrm{T}}\pi_\infty \\ -\pi_\infty^{\mathrm{T}}K_1K_1^{\mathrm{T}} & \pi_\infty^{\mathrm{T}}K_1K_1^{\mathrm{T}}\pi_\infty \end{bmatrix}. \quad (15)$$

For an ideal camera as suggested in Ref. [9], we have $u_0 = v_0 = 0$, $s = 0$ and $f_x = f_y \triangleq f_i$, which lead to the following camera matrix:

$$K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (16)$$

Thus, Eq. (15) can be simplified as [1]

$$\Omega^{*'}_{proj} = \begin{bmatrix} f_1^2 & 0 & 0 & -f_1^2\pi_{\infty_1} \\ 0 & f_1^2 & 0 & -f_1^2\pi_{\infty_2} \\ 0 & 0 & 1 & -\pi_{\infty_3} \\ -f_1^2\pi_{\infty_1} & -f_1^2\pi_{\infty_2} & -\pi_{\infty_3} & f_1^2\pi_{\infty_1}^2 + f_1^2\pi_{\infty_2}^2 + \pi_{\infty_3}^2 \end{bmatrix}. \quad (17)$$

Let

$$a = f_1^2, \quad b = -f_1^2\pi_{\infty_1}, \quad c = -f_1^2\pi_{\infty_2}, \quad d = -\pi_{\infty_3}, \quad e = f_1^2\pi_{\infty_1}^2 + f_1^2\pi_{\infty_2}^2 + \pi_{\infty_3}^2, \quad (18)$$

Eq. (14) becomes

$$\begin{bmatrix} f_i^2 & 0 & 0 \\ 0 & f_i^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \propto P_{proj_i}\begin{bmatrix} a & 0 & 0 & b \\ 0 & a & 0 & c \\ 0 & 0 & 1 & d \\ b & c & d & e \end{bmatrix}P^{\mathrm{T}}_{proj_i} \quad (19)$$

and we can obtain the following system of linear equations (see the Appendix):

$$\begin{aligned} k_{11}a + k_{12}b + k_{13}c + k_{14}d + k_{15}e + k_{16} &= 0, \\ k_{21}a + k_{22}b + k_{23}c + k_{24}d + k_{25}e + k_{26} &= 0, \\ k_{31}a + k_{32}b + k_{33}c + k_{34}d + k_{35}e + k_{36} &= 0, \\ k_{41}a + k_{42}b + k_{43}c + k_{44}d + k_{45}e + k_{46} &= 0. \end{aligned} \quad (20)$$

When only two images are available, instead of a unique solution, only a family of solutions can be determined for Eq. (20). Even if the rank 3 constraint for $\Omega^{*'}_{proj}$ in Eq. (17) is imposed, one still ends up with four possible solutions [9]. To overcome this difficulty, Pollefeys et al. [14] use additional scene constraint of vanishing points for a constant

[1] This matrix is essentially the same as Eq. (6) given in Ref. [9], except for some minor changes in notation.

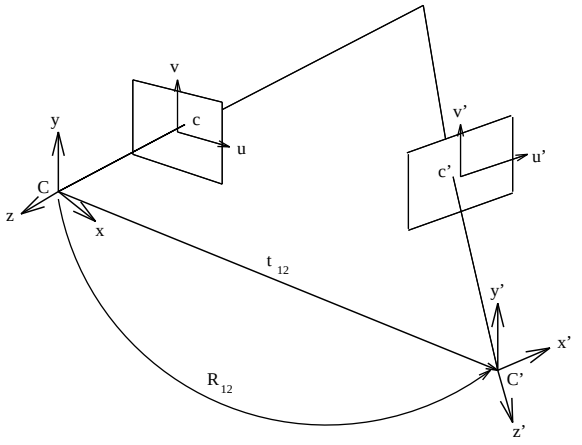

Fig. 1. Stereo head configuration.

camera to obtain a unique solution. In the next section, without additional scene constraints or more knowledge about the camera, we show that, not only for a constant camera but also for that with a varying focal length, it is possible to obtain a closed form solution using two images obtained from certain camera motions, or equivalently from special configurations of a stereo head.

4. Linear methods for stereo head

Consider the stereo configuration in which a pair of cameras are mounted on a lateral rig and are allowed to vary their angles of vergence freely, as shown in Fig. 1. Let R_{12} be the rotation from the first camera to the second one, and t_{12} be the corresponding translation vector. The above configuration is very similar to that in Ref. [18]; however, it leads to a more general vergence-based stereovision system.

Consider the self-calibration problem for such a general stereo head. Since only two images are involved, Eq. (8) becomes

$$H_{12}^{\infty} = H_{12} - e_r \pi_{\infty}^{\mathrm{T}} \propto K_2 R_{12} K_1^{-1}, \tag{21}$$

where e_r is the epipole in the second image, H_{12} denotes the homography, obtained here by $[e_r] \times F$ with F being the fundamental matrix relating these two images, and H_{12}^{∞} represents the corresponding infinity homography. Note that the last term of Eq. (21) is the expression of the infinity homography in the Euclidean space adopted in Ref. [19].

For a general rotation, Eq. (21) can be expressed as

$$\begin{bmatrix} h_{11} - e_{r_1}\pi_{\infty_1} & h_{12} - e_{r_1}\pi_{\infty_2} & h_{13} - e_{r_1}\pi_{\infty_3} \\ h_{21} - e_{r_2}\pi_{\infty_1} & h_{22} - e_{r_2}\pi_{\infty_2} & h_{23} - e_{r_2}\pi_{\infty_3} \\ h_{31} - e_{r_3}\pi_{\infty_1} & h_{32} - e_{r_3}\pi_{\infty_2} & h_{33} - e_{r_3}\pi_{\infty_3} \end{bmatrix}$$

$$\propto \begin{bmatrix} \frac{f_2}{f_1}(\cos\phi\cos\kappa) & \frac{f_2}{f_1}(\sin\omega\cos\phi\cos\kappa + \cos\omega\sin\kappa) & f_2(-\cos\omega\sin\phi\cos\kappa + \sin\omega\sin\kappa) \\ \frac{f_2}{f_1}(-\cos\phi\sin\kappa) & \frac{f_2}{f_1}(-\sin\omega\sin\phi\sin\kappa + \cos\omega\cos\kappa) & f_2(\cos\omega\sin\phi\sin\kappa + \sin\omega\cos\kappa) \\ \frac{f_2}{f_1}\sin\phi & \frac{f_2}{f_1}(-\sin\omega\cos\phi) & \cos\omega\cos\phi \end{bmatrix}, \tag{22}$$

where κ, ϕ and ω, denote the rotation-angle around the x-axis, the y-axis and the z-axis, respectively.

Based on Eq. (22), our aim in the following is to develop closed form expressions for the location of the plane at infinity (π_{∞_i}'s) for special stereo configurations which result in simpler expressions of R_{12}. Such special configurations are classified into two categories: the first class (Cases 1–3) involves rotations with respect to one of the coordinate axes while the second class (Case 4–6) covers rotations around two of the coordinate axes. Without loss of generality, the world frame is assumed to be aligned with the first camera; therefore R_{12} is simply denoted as R_2.

Case 1: $\omega \neq 0$, $\phi = \kappa = 0$. Here only rotation around the x-axis is involved and Eq. (22) becomes

$$\begin{bmatrix} h_{11} - e_{r_1}\pi_{\infty_1} & h_{12} - e_{r_1}\pi_{\infty_2} & h_{13} - e_{r_1}\pi_{\infty_3} \\ h_{21} - e_{r_2}\pi_{\infty_1} & h_{22} - e_{r_2}\pi_{\infty_2} & h_{23} - e_{r_2}\pi_{\infty_3} \\ h_{31} - e_{r_3}\pi_{\infty_1} & h_{32} - e_{r_3}\pi_{\infty_2} & h_{33} - e_{r_3}\pi_{\infty_3} \end{bmatrix}$$

$$\propto \begin{bmatrix} \frac{f_2}{f_1} & 0 & 0 \\ 0 & \frac{f_2}{f_1}\cos\omega & f_2\sin\omega \\ 0 & -\frac{\sin\omega}{f_1} & \cos\omega \end{bmatrix}. \tag{23}$$

From the four zeros on the right-hand side, we have

$$\pi_{\infty_1} = \frac{h_{21}}{e_{r_2}} \tag{24}$$

$$= \frac{h_{31}}{e_{r_3}}, \tag{25}$$

$$\pi_{\infty_2} = \frac{h_{12}}{e_{r_1}}, \tag{26}$$

$$\pi_{\infty_3} = \frac{h_{13}}{e_{r_1}}. \tag{27}$$

Thus, $\pi_{\infty} = [\pi_{\infty_1}\ \pi_{\infty_2}\ \pi_{\infty_3}]$ can be determined easily.[2] Given π_{∞}, the intrinsic camera parameters can be obtained

[2] In the implementation, two values of π_{∞_1} obtained from Eqs. (24) and (25), respectively, are found to be quite similar. The average of the two values are used as a robust estimation of π_{∞_1} in the experiments.

as follows. Since $d = -\pi_{\infty_3}$ in Eq. (18), one can obtain a (and f_1) by solving Eq. (20). On the other hand, for the second focal length f_2, it can be obtained from Eq. (19) directly. First, the scale factor in Eq. (19) can be calculated from the third diagonal element of the resulting 3×3 matrices as

$$\lambda = P^{(3)}_{proj_2} \Omega^{*'}_{proj} P^{(3)\mathrm{T}}_{proj_2}, \tag{28}$$

where $P^{(k)}_{proj_i}$ represents the kth row of P_{proj_i} in Eq. (3). Then, f_2 can be calculated from the first two diagonal elements of the matrices as

$$f_2 = \sqrt{\tfrac{1}{\lambda} P^{(1)}_{proj_2} \Omega^{*'}_{proj} P^{(1)\mathrm{T}}_{proj_2}} \tag{29}$$

or [3]

$$f_2 = \sqrt{\tfrac{1}{\lambda} P^{(2)}_{proj_2} \Omega^{*'}_{proj} P^{(2)\mathrm{T}}_{proj_2}} \tag{30}$$

For the 3D metric reconstruction, only the first camera matrix K_1 is needed in determining T_{PM} in Eq. (12) for the self-calibration procedure; therefore, the intrinsic parameters of the second camera are not essential for this purpose. Nevertheless, Eqs. (28)–(30) give a possible way of propagating of the calibration results from one image to another using the absolute dual quadric.

Case 2: $\phi \neq 0,\ \omega = \kappa = 0$. This corresponds to a rotation around the y-axis. In this case, Eq. (22) becomes

$$\begin{bmatrix} \times & h_{12} - e_{r_1}\pi_{\infty_2} & \times \\ h_{21} - e_{r_2}\pi_{\infty_1} & \times & h_{23} - e_{r_2}\pi_{\infty_3} \\ \times & h_{32} - e_{r_3}\pi_{\infty_2} & \times \end{bmatrix} \propto \begin{bmatrix} \times & 0 & \times \\ 0 & \times & 0 \\ \times & 0 & \times \end{bmatrix}, \tag{31}$$

where $\times$ denotes a value we do not care. Equations similar to Eqs. (24)–(27) can be established to find π_∞. Then, f_1 and f_2 can be obtained in exactly the same way as in Case 1.

Case 3: $\kappa \neq 0,\ |\omega| + |\phi| \approx 0$. In theory, for $\kappa \neq 0$ and $\omega = \phi = 0$, optical axes of the two cameras are parallel and hence results in a critical motion. [4] Therefore, we assume that at least a small rotation exists around one of x- or y-axis in this case that Eq. (22) can be expressed as

$$\begin{bmatrix} \times & \times & h_{13} - e_{r_1}\pi_{\infty_3} \\ \times & \times & h_{23} - e_{r_2}\pi_{\infty_3} \\ h_{31} - e_{r_3}\pi_{\infty_1} & h_{32} - e_{r_3}\pi_{\infty_2} & \times \end{bmatrix} \propto \begin{bmatrix} \times & \times & \approx 0 \\ \times & \times & \approx 0 \\ \approx 0 & \approx 0 & \times \end{bmatrix}. \tag{32}$$

However, because the scaling factor is unknown, using the four near zero elements, as in Cases 1 and 2, is not practical (and will lead to numerically unstable results). In order to obtain more reliable results, instead of using Eq. (32), consider the first two diagonal elements of the matrices:

$$\begin{bmatrix} h_{11} - e_{r_1}\pi_{\infty_1} & \times & \times \\ \times & h_{22} - e_{r_2}\pi_{\infty_2} & \times \\ \times & \times & \times \end{bmatrix} \propto \begin{bmatrix} \frac{f_2}{f_1}\cos\kappa & \times & \times \\ \times & \frac{f_2}{f_1}\cos\kappa & \times \\ \times & \times & \times \end{bmatrix} \tag{33}$$

and we have

$$\pi_{\infty_1} = \frac{h_{11} - h_{22} + e_{r_2}\pi_{\infty_2}}{e_{r_1}}. \tag{34}$$

Thus, b can be expressed as a function of a and c as

$$\begin{aligned} b &= -f_1^2 \pi_{\infty_1} \\ &= \frac{-h_{11} + h_{22}}{e_{r_1}} f_1^2 - \frac{e_{r_2}}{e_{r_1}} f_1^2 \pi_{\infty_2} \\ &\triangleq k_{b_1} a + k_{b_2} c. \end{aligned} \tag{35}$$

With Eqs. (20) and (35), the five variables defined in Eq. (18) and thus the four parameters, f_1, π_{∞_1}, π_{∞_2} and π_{∞_3}, can be solved.

Case 4: $\omega = 0,\ \phi \neq 0,\ \kappa \neq 0$. In this case, Eq. (22) becomes

$$\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & h_{32} - e_{r_3}\pi_{\infty_2} & \times \end{bmatrix} \propto \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & 0 & \times \end{bmatrix}. \tag{36}$$

Therefore, we have

$$\pi_{\infty_2} = \frac{h_{32}}{e_{r_3}}. \tag{37}$$

For the other three parameters, since $c = -a\pi_{\infty_2}$, one can obtain their solutions by solving Eq. (20), as in Case 1.

Case 5: $\phi = 0,\ \omega \neq 0,\ \kappa \neq 0$. Analogous to Case 4, we have

$$\pi_{\infty_1} = \frac{h_{31}}{e_{r_3}}. \tag{38}$$

[3] Similarly, two values of f_2 obtained from Eqs. (29) and (30), respectively, are averaged to give a robust estimation.

[4] The so-called critical motions correspond to the special camera motions wherein the calibration parameters cannot be determined uniquely by using the absolute conic, or the absolute dual quadric, alone. In the other cases, the critical motions are also possible. For instance, in Case 2, if t_{12} has zero component in the y direction, we also have a critical motion. For more details, one may refer to Ref. [20].

And the other parameters can be obtained similar to that done in Case 4.

Case 6: $\kappa = 0,\ \omega \neq 0,\ \phi \neq 0$. The calculations are similar to those done in Cases 4 and 5 and are omitted for brevity.

5. 3D metric reconstruction

An approach to self-calibration, as outlined in Refs. [12,14,17], includes a step-by-step procedure of projective reconstruction, affine reconstruction, and finally the metric reconstruction from multiple images. A similar approach but using only two images, according to the linear algorithm proposed in this paper can be summarized as follows.

The 3D metric reconstruction procedure

Stage 1: Find the two projective projection matrices:

$$P_{proj_1} = [I_{3\times 3}|0_3],$$

$$P_{proj_2} = [H_{12}|e_r]. \tag{39}$$

Stage 2: Obtain T_{PM}, i.e., find the four parameters f_1, π_{∞_1}, π_{∞_2} and π_{∞_3}.

Stage 3: Derive the projection matrix for the Euclidean space:

$$P_{euc_i} = P_{proj_i} T_{PM}^{-1}, \quad i = 1, 2. \tag{40}$$

Stage 4: Obtain the metric structure through the SVD-based 3D reconstruction method given in Ref. [21].

Note that in Stage 1, H_{12} is obtained by $[e_r] \times F$ with F being the fundamental matrix relating these two images.

6. Experiment results

In this section, calibration results obtained with the proposed linear methods are presented. For a self-calibration approach, the accuracy of the perspective reconstruction affects the results remarkably. The modified eight-point algorithm is adopted in the implementation, which usually gives a satisfactory fundamental matrix, F, even under noisy conditions [22]. Using the derived fundamental matrix, the metric reconstruction is then conducted.

The performance of the proposed approach is examined using both synthetic and real data. In the former, statistical results such as means and standard deviations are provided for f_1, f_2, π_∞ and the 3D reconstruction error, respectively, under different noise conditions. In the latter, on the other hand, we measure the parallelism and orthogonality of 3D structures obtained from the metric reconstruction.

6.1. Experiments using synthetic data

The simulations are carried out for the stereo configurations discussed in Section 4. For brevity, we conduct one simulation for each of the two classes, i.e., Cases 2 and 5, to demonstrate the effectiveness of the proposed approach.

For both simulations, a set of synthetic data consisting of 50 3D points is used, as shown in Fig. 2. These points are generated randomly in a $100 \times 100 \times 100$ cube centered at (100, 100, 100). For the first simulation (SIM 1), two images of this synthetic scene are generated using two Euclidean projection matrices, $P_{euc_1} = K_1[R_1|t_1]$ and $P_{euc_2} = K_2[R_2|t_2]$ with $K_1 = \mathrm{diag}[650, 650, 1]$, $R_1(\omega_1 = 0^\circ, \phi_1 = 0^\circ, \kappa_1 = 0^\circ)$, $t_1 = [0, 0, 0]$, $K_2 = \mathrm{diag}[750, 750, 1]$, $R_2(\omega_2 = 0^\circ,\ \phi_2 = 15^\circ,\ \kappa_2 = 0^\circ)$, and $t_2 = [10, 20, 10]$. The image points are contained in bounding boxes of size 1600×1400 pixels and 1200×900 pixels in the first and the second images, respectively. For the second simulation (SIM 2), the first image is the same as that in SIM 1, while the second image is generated using $P'_{euc_2} = K'_2[R'_2|t'_2]$ with $K'_2 = \mathrm{diag}[750, 750, 1]$, $R'_2(\omega'_2 = 18^\circ,\ \phi'_2 = 0^\circ, \kappa'_2 = 12^\circ)$, and $t'_2 = [15, 20, 15]$, and the points are contained in bounding boxes of size 800×900 pixels.

Fig. 3 shows the mean and standard deviation of the estimation of the first focal length, f_1, for zero mean Gaussian noise, with standard deviation ranging from 0 to 1 pixel, added to locations of image points. For each noise level, the statistics are obtained from a total of 100 trials. While their standard deviations have the general trend to increase with the noise level, the means of these estimations approximately give the correct value $f_1 = 650$. Similar results can also be observed for the estimations of the second focal length, $f_2 = 750$, as shown in Fig. 4.

To evaluate the accuracy of the estimated π_∞, the real π_∞ is first obtained with Eq. (21) using the real rotation R_2 and the fundamental matrix F obtained from a pair of noise-free images. The angular error is then defined as the angle between the estimated π_∞ and the real π_∞. Fig. 5 shows the mean and standard deviation of the angular error for various noise levels.

As for the 3D reconstruction errors, since real locations of all 3D points are given for the simulation, we can directly measure the average distance between the true 3D structure and the corresponding structure recovered in metric space. The distance measurement procedure can be described as follows. First, the two structures are moved to have their centroids located at the origin of the world coordinate system. Secondly, a size normalization operation for both structures is carried out. Thirdly, pointwise Euclidean distances between the two structures are calculated for corresponding 3D feature points. And finally, we average these distances to obtain an estimation of the 3D reconstruction error. Fig. 6 shows the mean value and the corresponding standard deviation of the 3D reconstruction error.

Similarly, Figs. 7–10 show the results of SIM 2. It can be seen easily that satisfactory results similar to those obtained for SIM 1 can be obtained. Thus, for the special stereo configurations considered in this paper, the proposed linear methods are indeed capable of solving the self-calibration problem, in terms of accuracy and robustness.

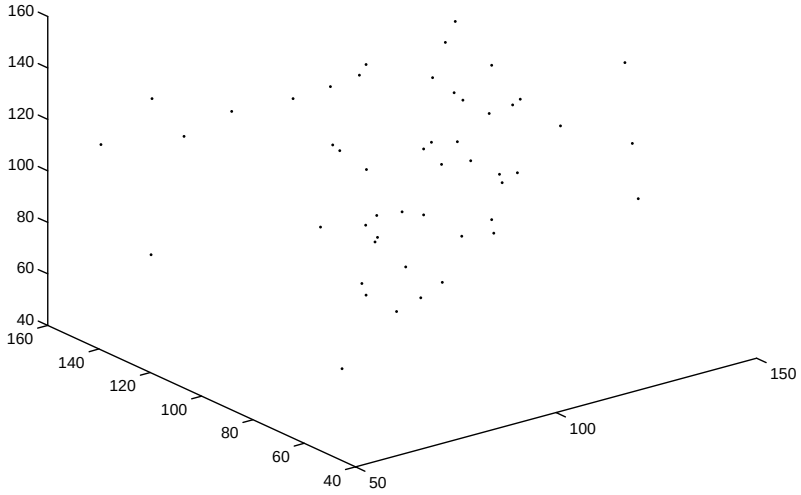

Fig. 2. The 3D structure for the simulation.

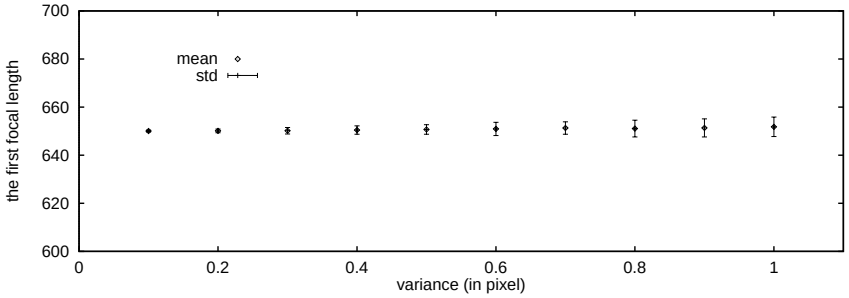

Fig. 3. The estimated first focal length (f_1) for various noise levels in SIM 1.

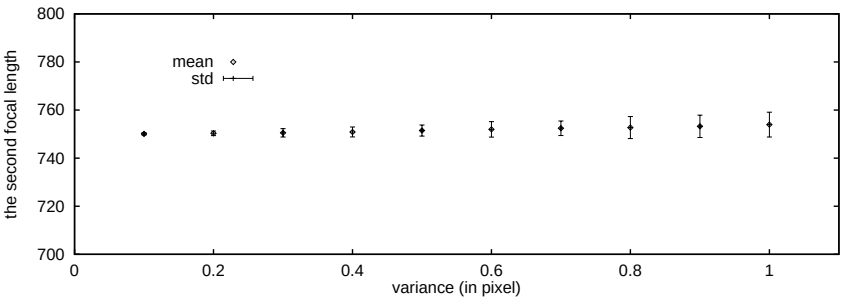

Fig. 4. The estimated second focal length (f_2) for various noise levels in SIM 1.

6.2. Experiments based on real images

In this subsection, reconstruction results based on the proposed linear methods for pairs of real images are presented. The images are obtained from the CMU image sequence [23]. The reason for using this sequence is simply because the underlying camera motions are essentially translation, which may approximately satisfy assumptions for the stereo configurations considered in Section 4. Fig. 11 shows a typical image with 36 ground truth points in this sequence. A subset of these point features is connected with 9 line segments (marked with A to I in Fig. 12) to facilitate the parallelism and orthogonality measurement of the recovered 3D metric structure. In the experiment, the principal points given in Ref. [23] are utilized as known intrinsic parameters in addition to the assumptions that $f_x/f_y = 1$ and $s = 0$.

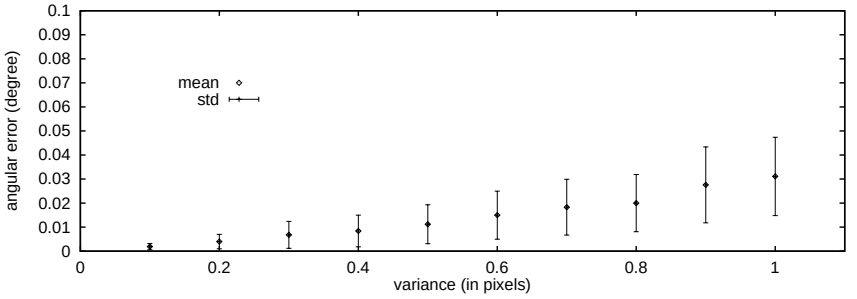

Fig. 5. Angular error for the estimates of π_∞ for various noise levels in SIM 1.

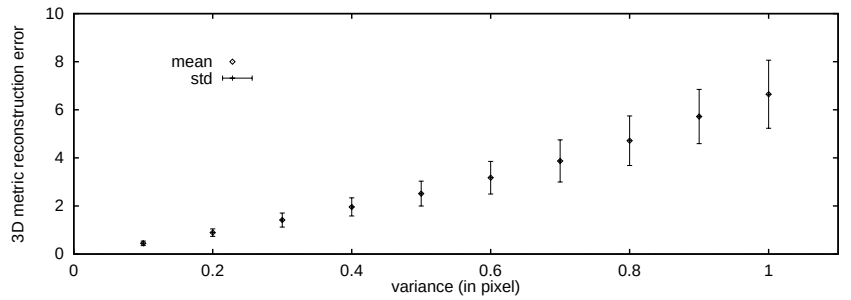

Fig. 6. The 3D metric reconstruction error for various noise levels in SIM 1.

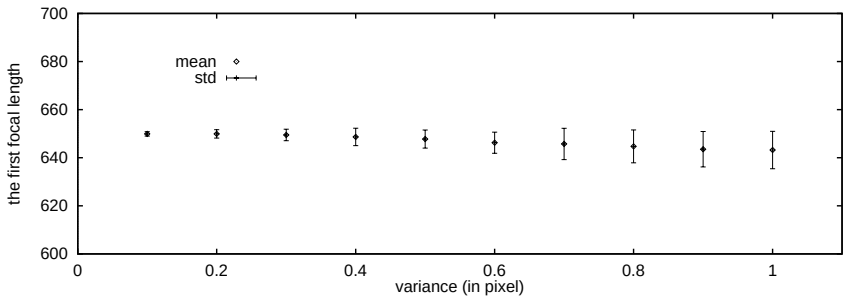

Fig. 7. The estimated first focal length (f_1) for various noise levels in SIM 2.

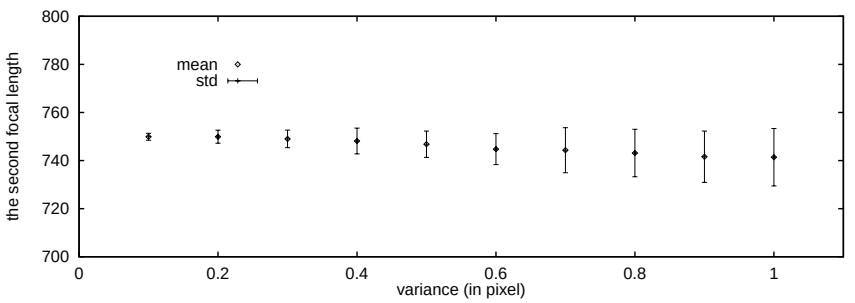

Fig. 8. The estimated second focal length (f_2) for various noise levels in SIM 2.

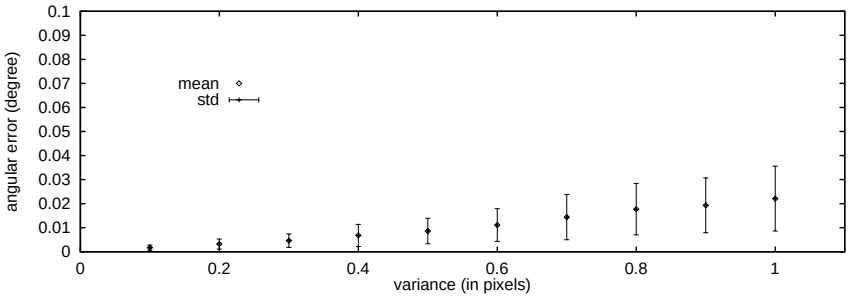

Fig. 9. Angular error for the estimates of π_∞ for various noise levels in SIM 2.

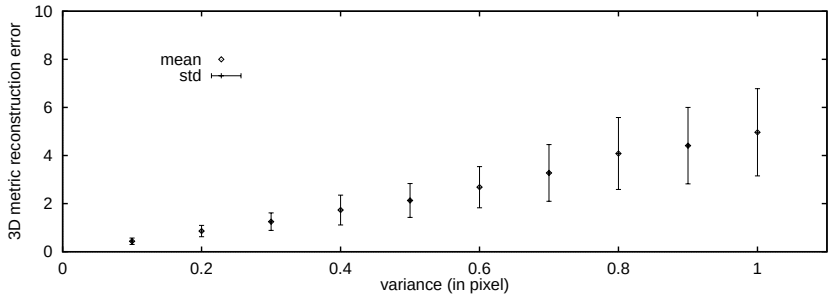

Fig. 10. The 3D metric reconstruction error for various noise levels in SIM 2.

Fig. 11. A typical image with 36 ground truth points in the CMU image sequence.

Accordingly, the proposed linear methods for the 3D metric reconstruction are applied to the first two images of the CMU image sequence.

Table 1 shows the true and estimated angles between 10 selected pairs of line segments for the recovered 3D structure. While μ represents the true angle between each pair of line segments, μ_1, μ_2, μ_3, μ_4, μ_5, and μ_6 denote the corresponding angles estimated according to Cases 1–6 discussed in Section 4, respectively. According to Table 1, the 3D structures recovered with the linear methods seem to preserve the parallelism and orthogonality of the original 3D structure satisfactorily for Cases 1, 2, 3 and 6 but not for

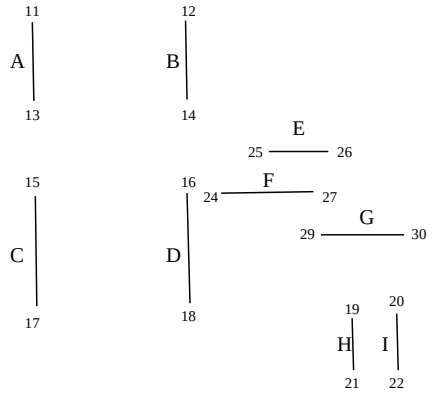

Fig. 12. Nine line-segments connecting a subset of point features.

Cases 4 and 5. In other words, assumptions for Cases 4 and 5 do not correspond to reasonable descriptions of camera orientations for obtaining the two images.

Since camera parameters are in principle nonlinearly encoded in the absolute dual quadric, to obtain more accurate calibration results, most of self-calibration algorithms utilize nonlinear cost functions to optimize their results using linear solutions as initial guesses. For example, according to Eq. (14), one can use the optimization criterion given in Ref. [17] to minimize

$$C'_F(K_i, \pi_\infty) = \sum_{i=2}^{N} ||F(K_i K_i^{\mathrm{T}}) - F(P_{proj_i} \Omega'^{*}_{proj} P_{proj_i})||_F, \tag{41}$$

where $F(A) \triangleq A/||A||_F$ with $||\cdot||_F$ denoting the Frobenius norm. Note that since only two images are used, N is equal to 2 in this case. It is readily observable from Table 2 that all of $\mu'_1, \ldots, \mu'_6$ obtained using $\mu_1, \ldots, \mu_6$ as initial guesses, respectively, give satisfactory angular estimates.

7. Conclusion

In this paper, we have shown that for some special configurations of a stereo head, certain linear equations resulting from the infinity homography can be added to a system of undetermined linear equations to solve the self-calibration problem. The stereo configurations considered here allow one camera to rotate around either one or two of the coordinate axes defined by another camera, possibly with different focal length. For such configurations, an approach to reconstructing 3D metric structures using only two images is then introduced. The performance of the proposed approach is verified with both synthetic and real data. Experiments with synthetic data show that satisfactory results can be obtained with the proposed linear methods alone. On the other hand, for real images obtained with more general stereo configurations, and possibly with measurement noises, results obtained from the linear methods may serve as reasonable initial guesses for the adopted nonlinear optimization procedure.

Appendix

The coefficients of the system of linear equations given in Eq. (20) are listed in detail in Table 3. [5]

Table 1
True and estimated angles (in degrees) between 10 selected pairs of line segments

Line-segment pair	μ	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
A–B	1.2071	0.942	3.796	0.787	19.801	1.370	0.785
C–D	0.9094	0.560	2.251	0.963	11.965	7.074	0.962
E–F	2.5379	1.081	1.112	1.106	2.239	1.923	1.097
F–G	1.8160	1.925	3.837	1.664	2.001	2.176	1.666
H–I	1.5674	1.029	1.028	0.486	2.371	19.376	0.485
A–E	88.8999	89.711	93.927	89.154	116.747	96.513	89.151
C–G	89.3059	89.059	93.731	89.062	114.687	97.011	89.064
D–F	90.1267	91.481	91.312	90.121	100.726	106.261	90.130
E–I	89.4330	94.057	87.716	89.041	89.523	163.113	89.077
G–H	90.5845	92.094	89.949	90.217	96.129	139.637	90.235
Error mean		1.1265	2.0807	0.4392	10.0399	17.9296	0.4350
Error std		1.3392	1.5207	0.4592	10.6640	24.3892	0.4630

[5] Herein, p_{ij}'s denote the elements of P_{proj_i}.

Table 2
True and estimated angles through nonlinear optimization

Line-segment pair	μ	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
A–B	1.2071	0.900	0.343	0.787	0.800	1.401	0.785
C–D	0.9094	0.599	0.604	0.958	0.371	0.839	0.954
E–F	2.5379	1.081	1.194	1.105	1.963	1.553	1.097
F–G	1.8160	1.926	0.327	1.655	1.016	2.110	1.651
H–I	1.5674	1.007	0.632	0.487	0.836	1.875	0.487
A–E	88.8999	89.847	88.194	89.155	88.671	90.080	89.151
C–G	89.3059	89.193	89.412	89.069	91.475	88.936	89.075
D–F	90.1267	91.658	90.053	90.123	90.840	91.875	90.135
E–I	89.4330	94.318	89.348	89.035	87.502	94.806	89.068
G–H	90.5845	92.333	90.123	90.216	91.297	93.018	90.232
Error mean		1.3221	0.6370	0.4405	0.8807	1.2956	0.4361
Error std		1.4298	0.5181	0.4595	0.6419	1.6257	0.4623

Table 3

$k_{11} = p_{11}^2 + p_{12}^2 - (p_{21}^2 + p_{22}^2)$	$k_{12} = 2(p_{11}p_{14} - p_{21}p_{24})$
$k_{13} = 2(p_{12}p_{14} - p_{22}p_{24})$	$k_{14} = 2(p_{13}p_{14} - p_{23}p_{24})$
$k_{15} = p_{14}^2 - p_{24}^2$	$k_{16} = p_{13}^2 - p_{23}^2$
$k_{21} = p_{11}p_{21} + p_{12}p_{22}$	$k_{22} = p_{14}p_{21} + p_{11}p_{22}$
$k_{23} = p_{14}p_{22} + p_{12}p_{24}$	$k_{24} = p_{14}p_{23} + p_{13}p_{24}$
$k_{25} = p_{14}p_{24}$	$k_{26} = p_{13}p_{23}$
$k_{31} = p_{11}p_{31} + p_{12}p_{32}$	$k_{32} = p_{14}p_{31} + p_{11}p_{32}$
$k_{33} = p_{14}p_{32} + p_{12}p_{34}$	$k_{34} = p_{14}p_{33} + p_{13}p_{34}$
$k_{35} = p_{14}p_{34}$	$k_{36} = p_{13}p_{33}$
$k_{41} = p_{21}p_{31} + p_{22}p_{32}$	$k_{42} = p_{24}p_{31} + p_{21}p_{32}$
$k_{43} = p_{24}p_{32} + p_{22}p_{34}$	$k_{44} = p_{24}p_{33} + p_{23}p_{34}$
$k_{45} = p_{24}p_{34}$	$k_{46} = p_{23}p_{33}$

References

[1] O.D. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig?, Proceedings of the ECCV, 1992, pp. 563–578.

[2] R.I. Hartley, Projective reconstruction and invariants from multiple images, IEEE Trans. Pattern Anal. Mach. Intell. 16 (10) (1994) 1036–1041.

[3] O.D. Faugeras, Q.-T. Luong, S.J. Maybank, Camera self-calibration: theory and experiments, Proceedings of the ECCV, 1992, pp. 321–334.

[4] M. Armstrong, A. Zisserman, R. Hartley, Self-calibration from image triplets, Proceedings of the ECCV, 1996, pp. 3–16.

[5] O.D. Faugeras, Stratification of 3D vision: projective, affine, and metric representation, J. Opt. Soc. Am. 12 (3) (1995) 465–484.

[6] R.I. Hartley, Euclidean reconstruction from multiple views, in: Applications of Invariance in Computer Vision, Lecture Notes in Computer Science, Vol. 825, Springer, Berlin, 1994, pp. 237–256.

[7] R.I. Hartley, Self-calibration from multiple views with a rotating camera, Proceedings of the ECCV, 1994, pp. 471–478.

[8] A. Heyden, K. Åaström, Euclidean reconstruction from constant intrinsic parameters, Proceedings of the ICPR, 1996, pp. 339–343.

[9] M. Pollefeys, L. Van Gool, Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters, Proceedings of the ICCV, 1998, pp. 90–95.

[10] L. Quan, B. Triggs, A unification of autocalibration methods, Proceedings of the ACCV, 2000, pp. 917–922.

[11] B. Triggs, Autocalibration and the absolute quadric, Proceedings of the CVPR, 1997, pp. 609–614.

[12] A. Zisserman, P. Beardsley, I. Reid, Metric calibration of a stereo rig, Proceedings of the IEEE Workshop Representation of Visual Scenes, 1995, pp. 93–100.

[13] M. Pollefeys, L. Van Gool, M. Proesmans, Euclidean 3D reconstruction from image sequences with variable focal lengths, Proceedings of the ECCV, 1996, pp. 31–42.

[14] M. Pollefeys, L. Van Gool, Stratified self-calibration with the modulus constraint, IEEE Trans. Pattern Anal. Mach. Intell. 21 (8) (1999) 707–724.

[15] L. Apapito, E. Hayman, I. Reid, Self-calibration of a rotating camera with varying intrinsic parameters, Proceedings of the British Machine Vision Conference, 1998.

[16] J. Semple, G. Kneebone, Algebraic Projective Geometry, Oxford University, Oxford, UK, 1979.

[17] M. Pollefeys, Self-calibration and metric 3D reconstruction from uncalibrated image sequences, Ph.D. Thesis, Katholieke Universtieit Leuven, Belgium, 1999.

[18] M.J. Brook, L. De Agapito, D.Q. Huynh, L. Baumela, Direct Methods for Self-Calibration of a Moving Stereo Head, Proceeding of the ECCV, 1996.

[19] Q.-T. Luong, T. Vieville, Canonic representations for the geometries of multiple projective views, Technique Report UCB/CSD-93-772, 1993.

[20] P. Sturm, Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length, Proceedings of the BMVC, 1999.

[21] C. Rothwell, O. Faugeras, G. Csurka, Different paths towards projective reconstruction, Europe–China Workshop of Invariance and Geometric Modelling, 1995.

[22] R.I. Hartley, In defense of eight-point algorithm, IEEE Trans. Pattern Anal. Mach. Intell. 19 (6) (1997) 580–593.

[23] CIL Stereo Dataset, http://www.cs.cmu.edu/afs/cs/project/cil/http:ftp/html/cil-ster.html.

About the Author—JAIN-SHING LIU was born in Taipei, Taiwan, in 1970. He received the Ph.D. degree in Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan, ROC, in 2001. Since August 2001, he has been on the faculty of the Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung, Taiwan. His research interests include 3D modeling, computer vision and image processing.

About the Author—JEN-HUI CHUANG (S'86-M'91) received the B.S. degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, ROC, in 1980, the M.S. degree in Electrical and Computer Engineering from the University of California at Santa Barbara in 1983, and the Ph.D. degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 1991.

Between 1983 and 1985, he was a Design and Development Engineer with the LSI Logic Corporation, Milpitas, CA. Between 1989 and 1991, he was a Research Assistant with the Robot Vision Laboratory, Beckman Institute for Advanced Science and Technology, University of Illinois, Champaign, IL. Since August 1991, he has been on the faculty of the Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan. His research interests include 3D modeling, computer vision, speech and image processing and VLSI systems.Dr. Chuang is a Member of the Tau Beta Pi Society.