This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

Geometry-Based Camera Calibration Using Five Point Correspondences from a Single Image

Hua-Tsung Chen, *Member, IEEE*

***Abstract*—As an essential step in many computer vision tasks, camera calibration has been studied extensively. In this paper, we propose a novel calibration technique that, based on geometric analysis, camera parameters can be estimated effectively and accurately from just one view of only five corresponding points. Our core contribution is the geometric analysis for deriving the basic equations to realize camera calibration from four coplanar corresponding points and a fifth non-coplanar one. The position, orientation, and focal length of a zooming camera can be estimated directly with unique solution. The estimated parameters are further optimized by the bundle adjustment technique. The proposed calibration method is examined and evaluated on both computer simulated data and real images. Experimental results confirm the validity of the proposed method that camera parameters can be estimated with sufficient accuracy using just five point correspondences from a single image, even in the presence of image noise.**

***Index Terms*—Camera calibration, calibration object, geometric derivation, sports video, single image.**

I. Introduction

A. Motivation

Deriving metric information from 2D images, camera calibration is often used as an essential early stage in computer vision. The task of camera calibration is to estimate the parameters which govern the relationship between the 2D image perceived by a camera and the 3D information of the imaged object [1]. Camera *extrinsic parameters* define the relative location and orientation of the camera in the real world, and *intrinsic parameters* describe the camera model. Providing the geometric mapping between 2D image positions and 3D real world coordinates, camera calibration is an indispensable step in a wide spectrum of applications, such as metrology [1], video surveillance [2], [3], augmented reality presentation [4], [5], virtual content insertion [6], 3D reconstruction [7], and so on. Since it would be more flexible to accomplish camera calibration in a single image with less corresponding points, we are motivated to explore the possibility of reducing the number of corresponding points and propose a novel calibration technique.

Manuscript received February 15, 2016; revised May 31, 2016; accepted July 14, 2016. This work was supported in part by projects ICTL -103-Q528, ATU-103-W958, MOST-103-2218-E-009-020, and MOST-105-2221-E-009-065.

The author is with the Information & Communications Technology Lab, National Chiao Tung University, Hsinchu, Taiwan (e-mail: huatsung@cs.nctu.edu.tw).

B. Related Work

Camera calibration has been extensively researched in photogrammetry and, more recently, in computer vision. According to the dimension of the calibration objects, calibration techniques can be mainly classified into the following categories:

1) *3D reference object-based calibration* is conducted by capturing a 3D reference object with precisely known geometry in 3D space [8]–[10] (to cite a few). Techniques in this category have the advantage of high accuracy and efficiency in estimating the intrinsic and extrinsic parameters [11]. Furthermore, just one image of the reference object is sufficient to accomplish calibration. However, it suffers from the disadvantage that in a multi-camera setup, a 3D reference object is hard to observe without self-occlusion by all cameras simultaneously.

2) *2D plane-based calibration* is performed by observing a planar pattern shown at a few (at least two) orientations [12]–[15] (to cite a few). In no need of the knowledge of the plane motion, this category of calibration is very flexible that the 2D plane pattern (e.g., a chessboard or paper on which a checker pattern is drawn) is easy to prepare and no special calibration setup is required. However, techniques in this category require capturing several images of calibration objects with fixed camera intrinsic parameters. Namely, 2D plane-based calibration cannot be accomplished from a single image, nor from a sequence of images involving camera zooming.

3) *1D linear object-based calibration* is a new trend first proposed by Zhang [16] and thereafter has attracted considerable attention [17]–[21]. Zhang's method uses a 1D object containing three collinear points with known distances and requires at least six observations of the 1D object rotating around a fixed point. Later on, Wu *et al.* [17] show that the rotating 1D calibrating object is in essence equivalent to a 2D planar calibration object, and also show that the calibration principle still holds when the 1D object undergoes a planar motion rather than rotating around a fixed point. Enhanced calibration methods with 1D object under general motions are also proposed [18], [19]. Compared with the 2D and 3D methods, 1D calibration object with simpler geometry is easier to construct. Furthermore, in a multi-camera setup the 1D calibration object can be captured without self-occlusion by all cameras simultaneously, but 2D or 3D objects are hard to meet this requirement. However, as pointed out by Wang *et al.* [20], the calibration accuracy still demands improvements due to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

noise-contaminated measurements. The degenerate cases of 1D object-based calibration are investigated by Hammarstedt *et al.* [21]. Similar to 2D methods, techniques in this category still suffer from the disadvantage that calibration cannot be accomplished from a single image, nor from a sequence of images involving camera zooming.

There are other calibration techniques using no calibration object. By moving a camera with fixed intrinsic parameters, *self-calibration* [22]–[25] (to cite a few) estimates camera parameters from feature points in a static image scene. Due to no calibration object, reliable feature points in multiple views are critical. Analyzing a single image, another direction of calibration is based on vanishing points [26]–[28] or vanishing lines [29]. Given this background, vanishing points/lines computed from static scene structures (e.g., buildings) are often exploited for camera calibration [7], [30]. A recent overview on this category of calibration can be found in [1].

In sports video analysis, camera calibration also plays a vital role, enabling a variety of applications, such as semantic/tactic analysis [31], [32], 3D ball trajectory reconstruction [33], [34], free viewpoint video synthesis [35], computer-aided refereeing [36], etc. Instead of setting up a calibration object, a *passive* object such as the court with a known 3D model is often exploited to compute the camera projection matrix for sports videos. One classical way is based on the detection of *corresponding points*—the reference points whose coordinates are known in both the 3D real world and the 2D image, such as intersection points of court lines and characteristic points on *court objects* (e.g., a net-post in tennis or badminton). Then the camera projection matrix is computed by solving a set of linear equations obtained from these corresponding points [31]–[34] (to cite a few). The projection matrix has 12 entries, and (ignore scale) 11 degrees of freedom, so it is necessary to have 11 equations to solve for the matrix. Since each point correspondence leads to two equations, at a minimal *six* corresponding (reference) points are required to compute the matrix. For a more thorough study, please refer to [1].

Using only five corresponding points, Miyagawa *et al.* [37] propose a calibration method to estimate the focal length, orientation and position of a zooming camera based on two orthogonal 1D objects, each with three collinear points, one of which is shared. Basic equations are derived from the five observed points to yield initial estimates of camera parameters, and the bundle adjustment technique [38] is adopted for parameter refinement.

In fact, there are still some other great *minimal* algorithms capable of estimating the camera parameters using even less (<5) point correspondences by solving systems of multivariate polynomials [39]–[42] (to cite a few). These minimal algorithms are usually executed in conjunction with RANSAC to filter out incorrect correspondence (outliers). Though fewer point correspondences are required, these minimal approaches encounter some limitations. First, the exhaustive linearization operations might deteriorate computation efficiency severely. For example, Bujnak *et al.* [40] propose a state-of-the-art method on the basis of the ratio of distance constraint and the polynomial solving techniques via Gröbner basis and hidden variable resultant. As mentioned by Sattler *et al.* [43], this method needs to perform Gauss-Jordan elimination on a 154 × 180 matrix followed by computing the eigenvalues of a 10 × 10 action matrix, demanding much computation time. Recently, the Gröbner basis solver is significantly accelerated in [44]. However, as pointed out by Zheng *et al.* [45], it requires expertise in the Gröbner basis technique. Moreover, it is also very challenging, especially for nonexperts, to trace the large resultant matrixes and to implement the complex linearization technology. On the other hand, estimates based on minimal data provide multiple solutions in general. For instance, in [40] the Gröbner basis solver and hidden variable solver yield 10 and 32 solutions, respectively. Though some constraints can be used to filter the solutions, the numerical stability is still limited, especially in the presence of noise. In this paper, we focus on the geometric analysis-based direct approach with unique solution. Thus, we do not address the minimal approaches with multiple solutions in detail nor compare with them.

C. Contribution

With the foregoing motivation and limitations of the existing works, we investigate the possibility of reducing the number of corresponding points and propose a novel camera calibration method requiring only one single image with just five point correspondences. A brief comparison of the previous works and our proposed method is given in Table I, listing the main advantages and limitations. Requiring less corresponding points, our proposed calibration approach can be realized with an easy-to-prepare calibration object. Although the proposed method may not be able to outperform the methods using multiple frame or more corresponding points in terms of accuracy, it provides a flexible way to estimate camera parameters with satisfying accuracy.

The most relevant previous work for this paper is the calibration method proposed by Miyagawa *et al.* [37], which also uses five corresponding points and estimates the focal length, orientation and position of a zooming camera. However, the errors of their estimated parameters increase rapidly with noise level. By comparison, our proposed method can yield more accurate estimation of camera parameters.

The main contributions and advantages of our work are summarized as follows.

(i) A novel geometric analysis for deriving the basic equations to realize camera calibration from four coplanar corresponding points and a fifth non-coplanar one is introduced. The position, orientation, and focal length[1] of a zooming camera can be estimated directly with unique solution.

(ii) The proposed method provides an alternative solution to some cases that other methods may fail to deal with, e.g., a single image or images from a single view (captured by a fixed camera) with insufficient visible reference points.

(iii) The proposed method is of high applicability in many types of scenes. For example, in an indoor scene, one will be

[1] Unlike other intrinsic camera parameters, which can be pre-calibrated in advance and are fixed in use, the focal length may change significantly and thus needs to be estimated from images together with the camera position and orientation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

TABLE I
COMPARISON OF PREVIOUS WORKS AND OUR PROPOSED METHOD.

Calibration method	Advantage	Limitation
3D calibration	● High accuracy ● Single image	● Self-occlusion of the calibration object ● At least 6 corresponding points
2D calibration	● High accuracy ● An easy-to-prepare calibration object used	● Multiple images with fixed camera intrinsic parameters required ● At least 4 corresponding points per image
1D calibration	● An easy-to-prepare calibration object used ● No self-occlusion problem	● Multiple images with fixed camera intrinsic parameters required ● Low robustness against noise
Self-calibration	● No calibration object required	● Several reliable feature points in multiple views required
Vanishing point/line-based method	● No calibration object required ● Single image	● Specific scene structures required
Passive object-based approach	● High accuracy ● Single image	● Self-occlusion of the reference object ● At least 6 corresponding points
Minimal algorithm	● Single image ● Less than 5 corresponding points	● Multiple solutions ● Low robustness against noise
Miyagawa's method [37]	● Single frame ● Only 5 corresponding points ● An easy-to-prepare calibration object used	● No estimation for intrinsic parameters ● Accuracy deterioration caused by lens distortion ● Low robustness against noise
Our proposed approach	● Single frame ● Only 5 corresponding points ● An easy-to-prepare calibration object used	● No estimation for intrinsic parameters ● Accuracy deterioration caused by lens distortion

able to set four reference points on the floor (e.g., corners of ceramic tiles) and a fifth one on a wall to conduct our calibration method. In an outdoor scene, one can simply set up a vertical pole, and then takes its top end and four pre-selected (or pre-marked) points on the ground as reference points. For applications of sports video analysis, it is also easy to find four planar points (e.g., intersections of court lines) and a non-planar one (e.g., a point on a court object). The proposed calibration method will be highly flexible in many applications.

The rest of this paper is organized as follows. In Section II, examples of the 5-point reference objects are shown, and the geometry definition of the camera position and rotation angles used in this paper is illustrated. In Section III, we elaborate the proposed geometric analysis and derive the equations for achieving camera calibration using only five corresponding points from a single image. In Section IV, we explain how to obtain the camera position for a special case. Then, we give a brief summary of the proposed camera calibration algorithm in Section V, and apply the proposed approach to sports videos in Section VI. Section VII presents experimental results of both computer simulated data and real images, and finally Section VIII concludes this paper and mentions future work.

II. EXAMPLES OF 5-POINT REFERENCE OBJECTS AND CAMERA MODEL

Fig. 1 presents two examples of the 5-point reference (calibration) objects that can be used in our camera calibration method. Each object contains four coplanar but non-collinear points: A, B, C, and D, and a fifth non-coplanar point E. The 3D real world positions and 2D image coordinates of the five reference points are known. In the 2D image coordinate system $p = (x, y)$, the five points corresponding to the reference points A, B, ..., E are denoted by $p_a = (x_a, y_a)$, $p_b = (x_b, y_b)$, ..., $p_e = (x_e, y_e)$, respectively. In fact, the reference object used in our camera calibration method is not restricted to the two in Fig. 1 only. Any four coplanar but non-collinear points and a fifth non-coplanar one can be utilized as a reference object for our

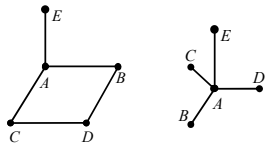


Fig. 1. Examples of 5-point reference objects.

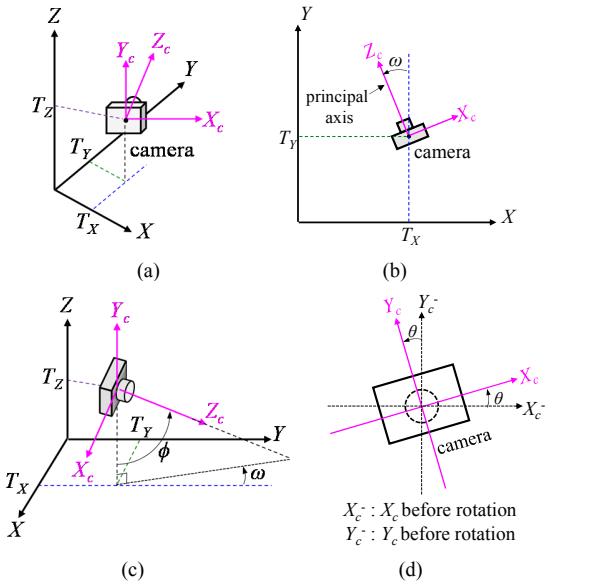


Fig. 2. Illustration of the camera position and rotation angles. (a) Camera position (T_X, T_Y, T_Z) in the world coordinate system XYZ. (b) Camera position on the XY-plane and rotation angle ω around the Y_c-axis. (c) Rotation angle ϕ around the X_c-axis. (d) Rotation angle θ around the Z_c-axis.

calibration method to estimate camera parameters.

The geometry definition of the camera position and rotation angles is illustrated in Fig. 2. The camera position, also known as the *optical center*, is denoted by (T_X, T_Y, T_Z) in the world coordinate system XYZ. Without loss of generality, we assume the reference points A, B, C, and D are on the XY-plane. In addition, we define the camera coordinate system $X_cY_cZ_c$, of which the origin is at the camera center and the *principal axis*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

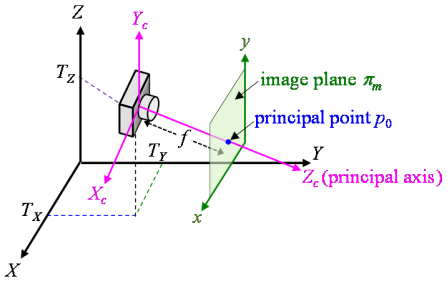


Fig. 3. Relationship among the world (X, Y, Z), camera (X_c, Y_c, Z_c), and image (x, y) coordinate systems.

(or principal ray) is in agreement with the Z_c-axis. The relationship among the world (X, Y, Z), camera (X_c, Y_c, Z_c), and image (x, y) coordinate systems is illustrated in Fig. 3.

A camera can be regarded as a mapping from the 3D real world to a 2D image. The principal camera of interest in this paper is *central projection*, and non-central cameras are out of the scope of this paper. The mapping relationship from a 3D world point $P = (X, Y, Z)$ to a 2D image point $p = (x, y)$ can be written in terms of matrix multiplication using homogeneous coordinates as:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{1}$$

where $\mathbf{P}$ is a 3×4 projection matrix with 11 degrees of freedom, defined up to an arbitrary scale. Furthermore, $\mathbf{P}$ can be factorized as:

$$\mathbf{P} = \mathbf{K}\mathbf{R}[\mathbf{I}\,|\,-\mathbf{T}], \tag{2}$$

where $\mathbf{I}$ is an identity matrix and $\mathbf{T}$ is the camera center $(T_X, T_Y, T_Z)^{\mathrm{T}}$. $\mathbf{R}$ is the rotation matrix describing the rotation around Y_c-, X_c-, and Z_c-axes in sequence[2] with respective rotation angles ω, ϕ, and θ, as:

$$\mathbf{R} = \begin{bmatrix} r_{00} & r_{01} & r_{02} \\ r_{10} & r_{11} & r_{12} \\ r_{20} & r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & \sin(\phi) \\ 0 & -\sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} \cos(\omega) & -\sin(\omega) & 0 \\ \sin(\omega) & \cos(\omega) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

$\mathbf{K}$ is an upper triangular matrix encoding the intrinsic parameters:

$$\mathbf{K} = \begin{bmatrix} f_x & s & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{4}$$

[2] Just to be clear: the rotation described here is first around the Y_c-axis, then around the *new* position of the X_c-axis, and finally around the *new* position of the Z_c-axis.

where f_x and f_y represent the *focal length* of the camera in terms of pixel dimensions in the x and y direction respectively. The parameter s is referred to as the *skew* parameter, which will be zero for most normal cameras, and (o_x, o_y) is the coordinate of the principal point (the point where the principal axis meets the image plane), termed p_0. In this paper, we assume $f_x = f_y = f$, $s = 0$, the principal point is at the projection center, and there is no lens distortion. Consequently, the transformation from the 3D real world to the 2D image plane can be formulated as:

$$x = f\frac{r_{00}(X - T_X) + r_{01}(Y - T_Y) + r_{02}(Z - T_Z)}{r_{20}(X - T_X) + r_{21}(Y - T_Y) + r_{22}(Z - T_Z)}, \tag{5}$$

$$y = f\frac{r_{10}(X - T_X) + r_{11}(Y - T_Y) + r_{12}(Z - T_Z)}{r_{20}(X - T_X) + r_{21}(Y - T_Y) + r_{22}(Z - T_Z)}. \tag{6}$$

III. Camera Parameter Estimation Using Five Point Correspondences

As mentioned in the previous section, the projection matrix $\mathbf{P}$ has 12 entries, and (ignore scale) 11 degrees of freedom, so it is necessary to have 11 equations to solve for $\mathbf{P}$. Since each point correspondence leads to two equations, at a minimal *six* such correspondences are required to solve for $\mathbf{P}$ (please refer to [1] for more details). In this section, we provide a geometric analysis-based approach and derive how to estimate camera parameters using only five point correspondences.

First of all, we consider the mapping between the world ground plane π_0 (the plane of $Z = 0$ in the XYZ coordinate system) and the image plane π_m. This projective transformation can be described by a plane-to-plane mapping (a homography) $\mathbf{H}$ as:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}. \tag{7}$$

The 3×3 homography transformation matrix $\mathbf{H}$ has 8 degrees of freedom, defined up to an arbitrary scale, so we need four point correspondences to estimate $\mathbf{H}$. We exploit the four coplanar points A, B, C, and D of the reference object to solve for $\mathbf{H}$ using the direct linear transform (DLT) algorithm [1].

Second, we compute the rotation angle θ using the *vanishing line*. It is known that the points at infinity in the 2D projective space form a line, which is usually called the *line at infinity* or vanishing line. Therefore, we can obtain the *horizon* (the vanishing line of the ground plane) from two points at infinity on the ground plane. From (7), we have:

$$x = \frac{h_{00}X + h_{01}Y + h_{02}}{h_{20}X + h_{21}Y + h_{22}}, \tag{8}$$

$$y = \frac{h_{10}X + h_{11}Y + h_{12}}{h_{20}X + h_{21}Y + h_{22}}. \tag{9}$$

Taking $X = \infty$ and $Y = \infty$ into (8) and (9), we get two vanishing points v_1 and v_2 on the image plane π_m:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

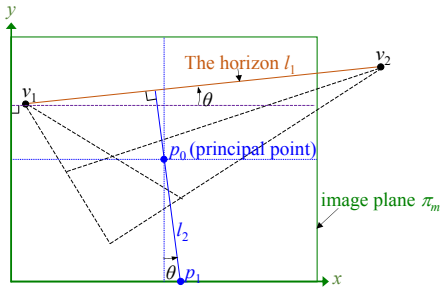


Fig. 4. Illustration of the horizon l_1 obtained from two vanishing points v_1 and v_2 on the image plane π_m, and the rotation angle θ. Perpendicular to l_1, l_2 goes through the principal point p_0 and intersects the bottom boundary of the image at p_1.

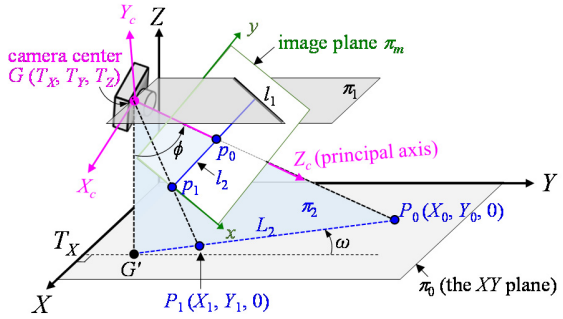


Fig. 5. Illustration for estimating the rotation angle ω. The plane π_1 goes through the camera center G and is parallel to the ground plane π_0. The plane π_2 perpendicular to both π_0 and π_1 contains G and G'. Points P_0 and P_1 are the projection of p_0 and p_1 on the ground plane.

$$v_1 = \left(\frac{h_{00}}{h_{20}}, \frac{h_{10}}{h_{20}}\right), v_2 = \left(\frac{h_{01}}{h_{21}}, \frac{h_{11}}{h_{21}}\right). \quad (10)$$

As illustrated in Fig. 4, the line l_1 passing through v_1 and v_2 is the horizon, and consequently we can estimate the rotation angle θ by

$$\theta = \tan^{-1}\left(\frac{\Delta y}{\Delta x}\right) = \tan^{-1}\left(\frac{\frac{h_{10}}{h_{20}} - \frac{h_{11}}{h_{21}}}{\frac{h_{00}}{h_{20}} - \frac{h_{01}}{h_{21}}}\right) = \tan^{-1}\left(\frac{h_{10}h_{21} - h_{11}h_{20}}{h_{00}h_{21} - h_{01}h_{20}}\right). \quad (11)$$

To compute the rotation angle ω, we have to acquire the projection of the principal axis on the ground plane first. As shown in Fig. 5, we can imagine a plane π_2 which contains both the camera center G and the principal point p_0, and is perpendicular to the ground plane π_0. This plane will also contain the vertical "footprint" of G on the ground plane, termed G'. Recalling the vanishing line formation (see pp. 216–217 in [1]), the vanishing line is constructed by intersecting the image plane π_m with a plane π_1 parallel to the ground plane π_0 through the camera center G. Therefore, π_2 will be perpendicular to π_1 too, and thereby the line l_2 formed by intersecting π_2 and π_m will be perpendicular to l_1.

Based on this inference, we draw a line perpendicular to l_1 and passing through p_0 in the image (i.e., l_2), as shown in Fig. 4, and obtain its intersection point p_1 with the bottom boundary of the image. Since **H** is a plane-to-plane homography mapping, we can project the image points p_0 and p_1 onto the ground plane π_0, termed P_0 and P_1, by the inverse form of (7), i.e.:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \mathbf{H}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (12)$$

As shown in Fig. 5, the line L_2 formed by connecting P_0 and P_1, which is the projection of the principal axis on the ground plane, will pass through G'. Let $P_0 = (X_0, Y_0, 0)$ and $P_1 = (X_1, Y_1, 0)$. Then, the rotation angle ω can be computed by

$$\omega = \tan^{-1}\left(\frac{X_1 - X_0}{Y_0 - Y_1}\right). \quad (13)$$

The next step is to compute the position of camera center $G = (T_X, T_Y, T_Z)$ using the non-coplanar reference point $E = (X_E, Y_E, Z_E)$, which corresponds to the image point p_e. As illustrated in Fig. 6, we project the image point p_e onto the ground plane at $P_e = (X_e, Y_e, 0)$ by (12). Let $E' = (X_E, Y_E, 0)$ and $G' = (T_X, T_Y, 0)$ be the vertical footprints of E and G on the ground plane, respectively. According to the pinhole camera model, the line L_e formed by connecting P_e and E will pass through the camera center G. Consequently, the line L_3 formed by connecting P_e and E' will pass through G'. Now we have two lines passing through G': L_2 and L_3. Fig. 7 presents a top-view illustration for the camera position and the lines L_2 and L_3 on the XY-plane. The coordinates of G' can be obtained by intersecting L_2 and L_3 using the following equations:

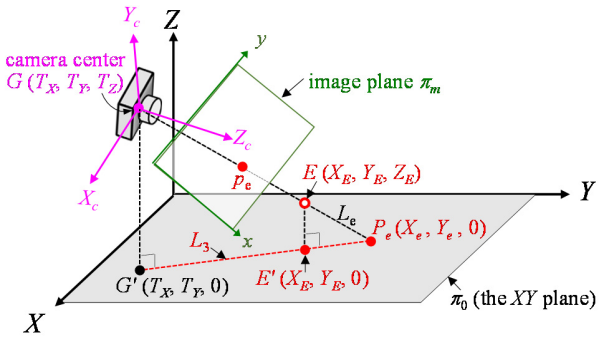


Fig. 6. Projecting the image point p_e onto the ground plane at P_e.

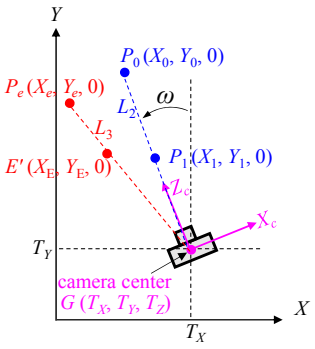


Fig. 7. Computing the camera position on the ground plane by intersecting L_2 and L_3.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

$$m_2 = \frac{Y_1 - Y_0}{X_1 - X_0},\ m_3 = \frac{Y_E - Y_e}{X_E - X_e}, \tag{14}$$

$$T_X = \frac{m_2 X_0 - m_3 X_e - Y_0 + Y_e}{m_2 - m_3}, \tag{15}$$

$$T_Y = m_2 (T_X - X_0) + Y_0, \tag{16}$$

$$\text{or } T_Y = m_3 (T_X - X_e) + Y_e, \tag{17}$$

where m_2 and m_3 are slopes of L_2 and L_3, respectively. However, m_2 (or m_3) cannot be obtained when $X_1 = X_0$ (or $X_E = X_e$). Thus, if $X_E = X_e$, we set $T_X = X_E$ and compute T_Y using (16), and if $X_1 = X_0$, we set $T_X = X_1$ and compute T_Y using (17). Since it is possible that L_2 is identical with L_3 (causing $m_2 = m_3$) when p_e is on the line l_2, we will derive the basic equations to obtain the camera position for this special case in the next section. If m_2 is not equal to m_3, we can obtain T_X and T_Y, and then compute the coordinate T_Z using the two similar triangles, $\Delta P_e E E'$ and $\Delta P_e G G'$ (see Fig. 6), as:

$$T_Z = Z_E \frac{|\overline{P_e G'}|}{|\overline{P_e E'}|}. \tag{18}$$

With the camera position $G = (T_X, T_Y, T_Z)$ and the point $P_0 = (X_0, Y_0, 0)$, where the principal axis meets the ground plane, being obtained (see Fig. 5), the angle ϕ can be computed by:

$$\phi = \tan^{-1}\left(\frac{\sqrt{(X_0 - T_X)^2 + (Y_0 - T_Y)^2}}{T_Z} \right). \tag{19}$$

From the obtained rotation angles θ, ϕ, and ω, we compute the elements r_{ij} $(i, j = 0, 1, 2)$ of the rotation matrix $\mathbf{R}$ as in (3). Then, taking the 3D and 2D coordinates of the five corresponding points A, B, ..., E into (5) and (6) yields a set of similar relations, and the focal length f can be solved from these similar relations using the DLT algorithm.

IV. Camera Position Estimation for The Special Case

As mentioned in the previous section, if p_e lies on l_2, L_2 will be in agreement with L_3. In such a case, the camera position $G = (T_X, T_Y, T_Z)$ cannot be computed using (14)–(18). In this section, we derive how to obtain the camera position for this special case.

On the image plane π_m, we first draw an auxiliary line l_4 parallel to l_2 (but not identical to l_2), and choose two arbitrary points p_2 and p_3 on l_4, as shown in Fig. 8. Let π_3 denote the plane containing the points G, p_2, and p_3. As illustrated in Fig. 9, the line L_4 formed by intersecting π_2 and π_3 will go through G (since G is a common point of π_2 and π_3) and will be parallel to π_m (since L_4 is parallel to both l_2 and l_4). As a result, we derive that L_4 is perpendicular to the principal axis. Let P_2 and P_3 be the 3D world points projected from p_2 and p_3 onto π_0, respectively, and L_5 denote the line passing through P_2 and P_3. After computing the intersection point of L_2 and L_5, termed K, we know the coordinates of K, E, E', P_e, and P_0 on the plane π_2,

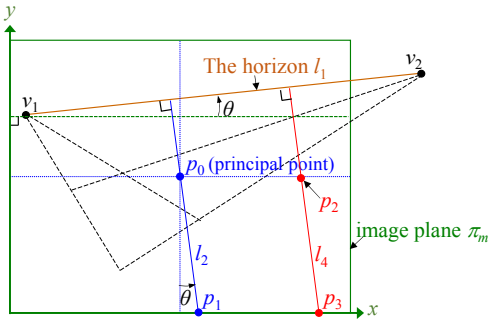


Fig. 8. Drawing an auxiliary line l_4 parallel to l_2 on the image plane π_m, and choosing two arbitrary points p_2 and p_3 on l_4.

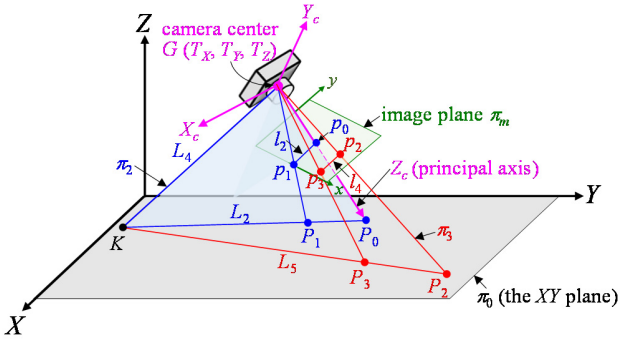


Fig. 9. Projecting the image points p_2 and p_3 onto the ground plane π_0 at P_2 and P_3, forming the line L_5, and computing the intersection point K of L_2 and L_5.

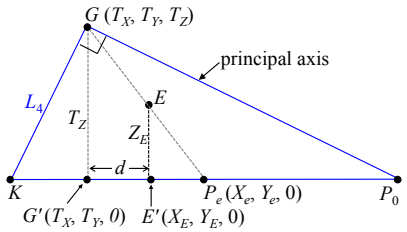


Fig. 10. Illustration of the feature points on the plane π_2 for estimating the camera position G.

as shown in Fig. 10. Now we attempt to compute the coordinates of G. For simplicity, the unknown distance between points K' and G' is termed d in the following equations. Since L_4 is perpendicular to the principal axis, we have:

$$|\overline{KP_0}|^2 = |\overline{KG}|^2 + |\overline{GP_0}|^2. \tag{20}$$

On the other hand, we have:

$$|\overline{KG}|^2 = |\overline{KG'}|^2 + T_Z^2,\ |\overline{GP_0}|^2 = |\overline{G'P_0}|^2 + T_Z^2. \tag{21}$$

From (20) and (21), we obtain:

$$\begin{aligned} |\overline{KP_0}|^2 &= |\overline{KG'}|^2 + T_Z^2 + |\overline{G'P_0}|^2 + T_Z^2 \\ &= 2T_Z^2 + (|\overline{KE'}| - d)^2 + (|\overline{E'P_0}| + d)^2 \end{aligned} \tag{22}$$

From the two similar triangles $\Delta P_e G G'$ and $\Delta P_e E E'$, we get:

$$\frac{T_Z}{Z_E} = \frac{d + |\overline{E'P_e}|}{|\overline{E'P_e}|},\ d = |\overline{E'P_e}| \left(\frac{T_Z}{Z_E} - 1 \right). \tag{23}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

Taking (23) into (22), we can solve T_Z. Again, from the two similar triangles $\Delta P_e G G'$ and $\Delta P_e E E'$, we can compute T_X and T_Y by:

$$T_X = X_e + (X_E - X_e)\frac{T_Z}{Z_E},\ T_Y = Y_e + (Y_E - Y_e)\frac{T_Z}{Z_E}. \quad (24)$$

With the camera position $G = (T_X, T_Y, T_Z)$ obtained, the rotation angle ϕ and the focal length f can be computed in the same way as described in the previous section.

V. Implementation

We summarize the proposed calibration method for camera parameter estimation based on the above derived equations as follows.

1) Locate on the captured image the five 2D points p_a, p_b, p_c, p_d, and p_e corresponding to the reference points A, B, C, D, and E.
2) Compute the homography transform **H** in (7) using the four coplanar reference points A, B, C, and D.
3) Derive the horizon l_1, of the image, through two vanishing points v_1 and v_2 by (10), and then compute the rotation angle θ using (11).
4) Locate two 3D points $P_0 = (X_0, Y_0, 0)$ and $P_1 = (X_1, Y_1, 0)$ by projecting the 2D image points p_0 and p_1 onto the world ground plane using (12), where p_0 is the principal point and p_1 is the intersection point of l_2 (the line perpendicular to l_1 and passing through p_0) and the image bottom boundary. Then, compute the rotation angle ω using (13).
5) If p_e does not lie on l_2, compute the camera position $G = (T_X, T_Y, T_Z)$ utilizing the reference point E by (14)–(18); otherwise, compute the camera position by the algorithm described in Section IV.
6) Compute the rotation angle ϕ by (19).
7) Compute the focal length f via solving a set of similar relations, yielded by taking the 3D and 2D coordinates of the five corresponding points into (5) and (6), using the DLT algorithm.

Though the proposed calibration approach can provide a fairly good initial estimate of camera parameters, the bundle adjustment technique [38], as adopted in [37], can be further applied in order to ensure the robustness against image noise. All the camera parameters are optimized using the Levenberg–Marquardt algorithm in the following equation:

$$\sum_{k=a,b,\ldots,e} \left\| p_k - p_k'(\theta, \omega, \phi, T_X, T_Y, T_Z, f) \right\| \rightarrow min. \quad (25)$$

where p_k' $(k = a, b, ..., e)$ are the 2D coordinates computed from the five reference points A, B, ..., E with the estimated camera parameters: θ, ω, ϕ, T_X, T_Y, T_Z, and f. As a result, all the parameters can be well refined.

VI. Application to Sports Videos

In sports video analysis, the topic of camera calibration attracts considerable attention. Many applications, such as tactic analysis [31], [32], computer-aided refereeing [36], mixed or augmented reality presentation [4]–[6], require camera calibration to derive the coordinate mapping between the 2D image and 3D real world. In this section, we demonstrate how to apply the proposed calibration method to sports videos and what advantage it has over the traditional 3D calibration method by reducing the number of corresponding points to five.

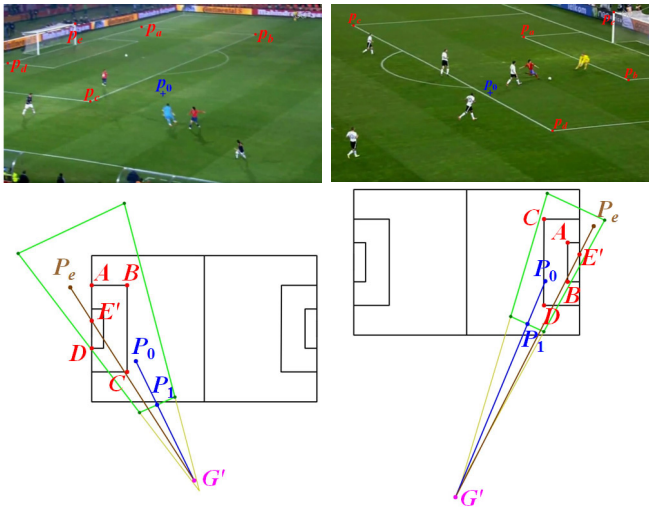
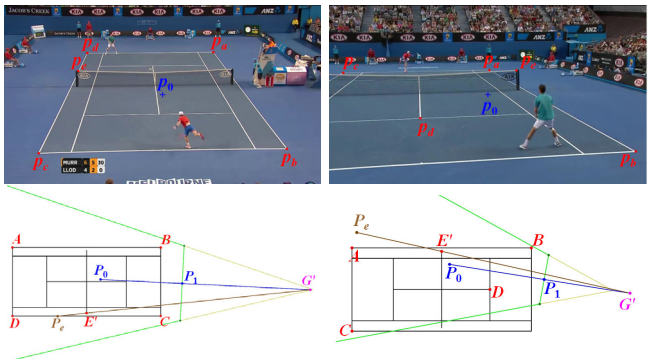


Fig. 11. Camera calibration in soccer videos.

Five popular sports are selected to test and verify the proposed calibration method, including soccer, tennis, volleyball, badminton, and basketball. For these sports, the four coplanar corresponding points are selected on the intersection points of court lines. As for the fifth point, i.e., the non-coplanar one, we locate it on a characteristic object: the top point of a goal-post in soccer, the top point of a net-post in tennis and badminton, a corner of the court-net in volleyball, and a corner of the backboard in basketball. For more details about the extraction of court lines and the detection of feature points on court objects, please refer to [4]–[6], [32]–[34], [46], etc.

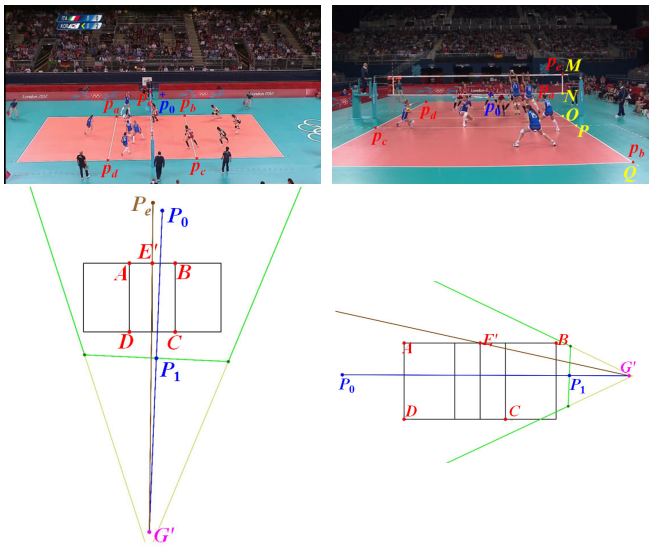
With the known 3D court model, camera parameters can be estimated by our proposed calibration method using the five extracted corresponding points, of which both 2D and 3D coordinates are known. Some calibration results of our proposed approach are demonstrated in Figs. 11–15. Take Fig. 11 for example to explain in more detail. The top row shows the images captured from different cameras in a soccer game, wherein five corresponding points are marked by red points (four court line intersections p_a, ..., p_d and one top point of a goal-post p_e) and the blue cross indicates the principal point p_0. The bottom row presents the corresponding bird-view images of the court, visualizing the geometric relations among the feature points: points A, ..., E corresponding to the image point p_a, ..., p_e, and points P_0, P_1, and P_e obtained by projecting the principal point p_0 and the image points p_1 and p_e onto the world ground using (12). (Please refer to Fig. 4 and Fig. 5 to recall the acquisition of P_0, P_1, p_0, and p_1.) The intersection point G' of lines P_eE' and P_0P_1 is the estimated ground position of the camera. With such an enriched illustration, one can clearly

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology



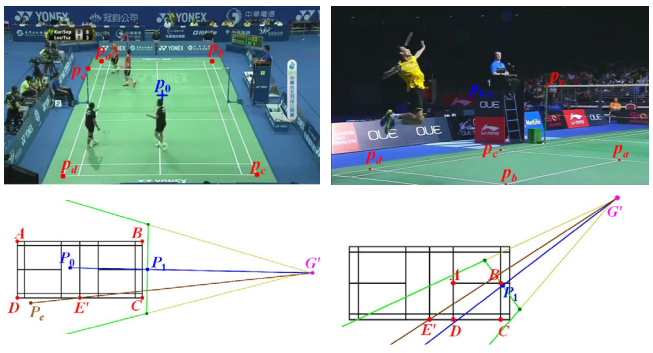
(a) (b)

Fig. 12. Camera calibration in tennis videos.



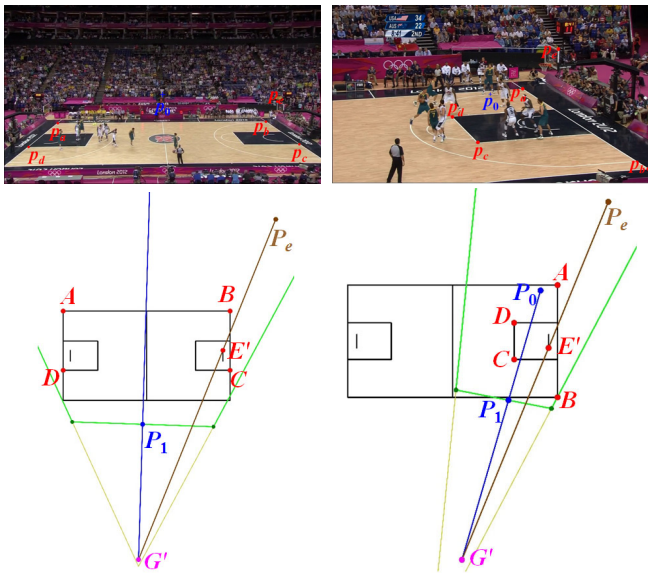
(a) (b)

Fig. 13. Camera calibration in volleyball videos. (M, N, O, P, and Q can be used for L-callib, where $M = P_e$ and $Q = P_b$.)



(a) (b)

Fig. 14. Camera calibration in badminton videos.



(a) (b)

Fig. 15. Camera calibration in basketball videos.

comprehend the camera location G' in the court, the viewing direction of the camera (from G' to P_0), and what region of the court is currently focused (the area bounded by green lines). Much more information can be further obtained from these cues to infer, for example, where an important event is taking place.

One significant advantage of our approach over other 3D calibration methods, which require at least 6 corresponding points, is that the 3D calibration methods will fail to treat the case where only one non-coplanar point is visible in the image due to the limitation of the camera viewing field or object occlusion. However, our proposed calibration method can still perform well, as demonstrated in Figs. 11(b), 12(b), and 14(b), wherein only one goal-post or net-post is visible in the images.

VII. Experimental Results

The proposed calibration method is examined and evaluated on both computer simulated data and real images. For comparison, we also faithfully implement a typical 3D object-based calibration method using eight corresponding points [1] and Miyagawa's approach [37]. Miyagawa's approach uses five points on two orthogonal 1D objects in L-, Γ-, or T- type. The two 1D objects, each of which has three collinear points, cross at one shared point and are perpendicular to each other. In our experiments the L-type reference object is used. In the following, the 3D object-based method using eight point correspondences, Miyagawa's approach using L-type reference object, and our proposed method using five point correspondences are respectively abbreviated to "3D-calib," "L-calib," and "5P-calib" for short.

A. Computer Simulations

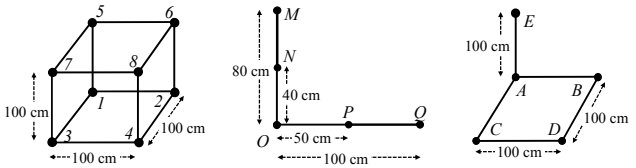
In this subsection, computer simulations are conducted to evaluate the accuracy of the existing methods and our proposed method. The image resolution is 1600 × 900. The intrinsic parameters of the simulated camera are set as follows: focal length f = 1200, aspect ratio 1, zero skew, principal point at (800, 450), and zero lens distortion. Table II presents the extrinsic parameters used in the simulation.

The reference object used for 3D-calib is a cube of edge length 100 cm, as shown in Fig. 16(a). To carry out L-calib, we simulate a virtual L-type object, as shown in Fig. 16(b), which consists of two perpendicular sticks crossing at a point O. Three points M, N, and O are set on one stick, and three points O, P, and Q are set on the other. As used in [37], the points M, N, P,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

TABLE II
CAMERA EXTRINSIC PARAMETERS USED IN THE SIMULATION.

Translation (cm)			Rotation (deg)		
T_X	T_Y	T_Z	ω	ϕ	θ
200	300	180	45	65	-5



(a) (b) (c)

Fig. 16. Reference objects used for (a) 3D-calib, (b) L-calib, and (c) 5P-calib.

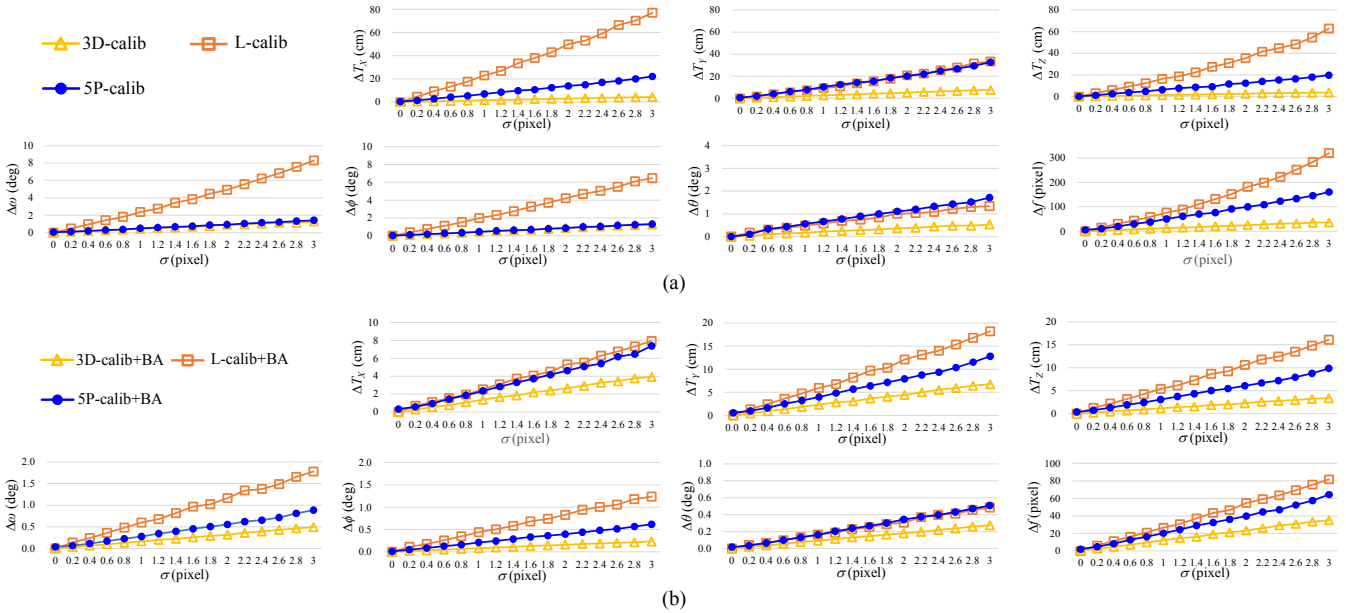


Fig. 17. Comparison of 3D object-based calibration (3D-calib), Miyagawa's approach (L-calib) [37], and our proposed calibration method using five point correspondences (5P-calib). (a)/(b): Parameter estimation errors without/with bundle adjustment.

and Q have the distances of 80 cm, 40 cm, 50 cm, and 100 cm from the point O, respectively. The one used for our proposed 5P-calib is, as shown in Fig. 16(c), a square $ABCD$ of side length 100 cm and a point E on a stick perpendicular to the square at the point A in a distance of 100 cm.

The reference points examined for each calibration approach are projected onto image planes according to the camera parameters listed in Table II. Furthermore, Gaussian noise with 0 mean and σ standard variation is added to the projected 2D points. We vary the noise level σ from 0.0 to 3.0 pixels, and conduct 1000 independent trials for each noise level. Finally, the absolute average errors between the ground truth and the camera parameters estimated by each approach are measured.

The comparison results of the existing methods and our approach are presented in Fig. 17, wherein the parameter estimation errors: ΔT_X, ΔT_Y, ΔT_Z, $\Delta\omega$, $\Delta\phi$, $\Delta\theta$, and Δf without and with bundle adjustment (abbr. BA) are given in Figs. 17(a) and 17(b), respectively. It is readily observable that 3D-calib, which uses more reference points than the other methods, yields very low errors, showing its good robustness against image noise. As for L-calib, the errors of the estimated parameters increase more rapidly with the noise level than the other methods. As explained in [37], such performance degradation is mainly because L-calib uses just five reference points. However, using the same number of points as L-calib, the proposed 5P-calib approach can estimate camera parameters with lower errors. Without BA, the errors of the camera rotation angles ω, ϕ, and θ estimated by 5P-calib are less than 2 degrees, and can further be reduced to less than 1 degree after applying BA refinement. For the estimated camera position, the relative errors with respective to the distance between the camera and the reference object are no more than 6%. The relative errors of the focal length without and with BA refinement are 13.3% (160.49/1200) and 5.3% (64.34/1200), respectively.

On the other hand, we also conduct experiments to compare the robustness against image noise among different reference points for 3D-calib, L-calib, and 5P-calib. Each time one reference point is chosen that its projected 2D point is added with higher-level Gaussian noise (σ = 4.0 pixels), and the others are added with lower-level noise (σ = 1.0 pixel). Five reference points are chosen in turn and 1000 independent trials are performed for each chosen point. The absolute average errors between the ground truth and the camera parameters estimated by 3D-calib, L-calib, and our 5P-calib (with BA refinement) are measured, as presented in Fig. 18. The experimental results show that 3D-calib, which uses eight reference points, is almost not affected by a single high-noised point. For L-calib, the reference point P seems more sensitive to image noise, compared with the other points. As for the proposed 5P-calib, one can reasonably infer that the reference point E may be more sensitive to image noise since it is the only

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

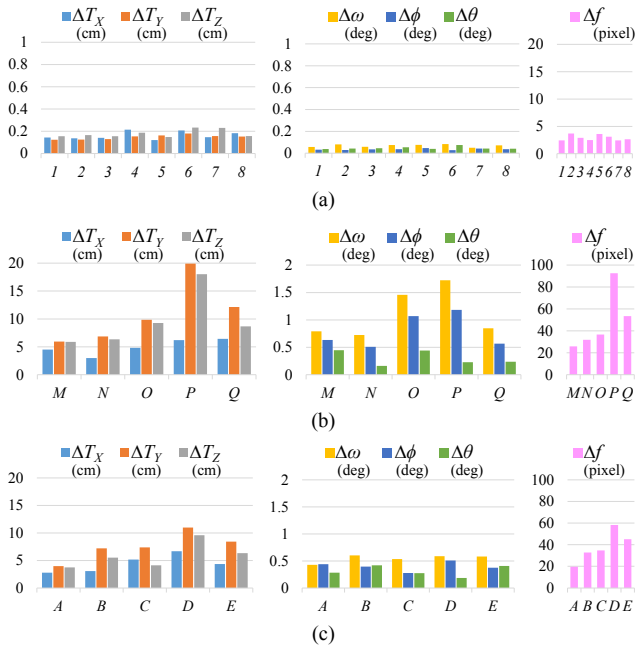


Fig. 18. Results of comparing the robustness against image noise among different reference points for (a) 3D-calib+BA, (b) L-calib+BA, and (c) 5P-calib+BA.

non-coplanar reference point. However, the results demonstrate that the point *E* show similar noise robustness to the other ones.

B. Real Images

To evaluate the accuracy of our proposed method on real data, we capture several images (with the ground-truth camera parameters associated) in different indoor and outdoor scenes, including a building hall, a courtyard, and a volleyball court. For the scenes of the building hall and courtyard, we use an off-the-shelf iron rack as a reference object to conduct 3D-calib, L-calib, and 5P-calib simultaneously. As shown in Fig. 19, points #0, #2, #3, #4, #5, #6, #7, and #8 of the iron rack can be used for 3D-calib, points #0, #1, #2, #5, and #9 for L-calib, and points #0, #2, #3, #4, and #5 for 5P-calib. For a volleyball court, the intersections of court lines and the corners of the net can be utilized. Figs. 20(a), (b), and (c) show the test images captured in a building hall, a courtyard, and a volleyball court, respectively.

In each scene, we capture thirty images from one camera position with five ω values, two ϕ values, and three θ values. The camera extrinsic parameters and image resolution used in each scene are listed in Table III. The reference points in each image are located manually. Table IV presents the absolute average errors measured between the ground truth and the camera parameters estimated by each method (with BA refinement).

Tested with real images, 3D-calib+BA performs very well for the three scenes that its estimation errors are the lowest for almost all the parameters. The cases that 3D-calib+BA yields slightly higher errors than other methods, e.g., $\Delta\phi$ in the courtyard scene and Δf in the volleyball court scene, may be due to the noise-contaminated measurement of the additional reference points. Both L-calib+BA and 5P-calib+BA can accurately estimate the rotation angles with the errors $\Delta\omega$ and $\Delta\phi$ around 1 to 3 degrees, and $\Delta\theta$ no more than 1 degree. However, L-calib+BA yields much larger errors in estimating the camera position.

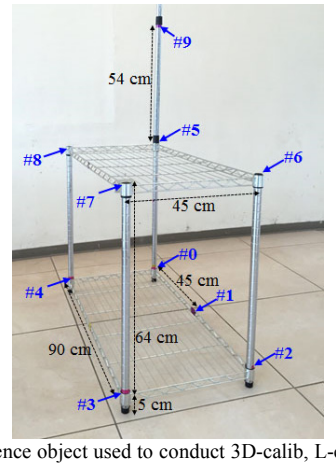


Fig. 19. The reference object used to conduct 3D-calib, L-calib, and 5P-calib simultaneously.

(a)

(b)

(c)

Fig. 20. Test images captured in (a) a building hall, (b) a courtyard, and (c) a volleyball court.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

TABLE III
CAMERA PARAMETERS AND IMAGE RESOLUTION USED IN EACH SCENE.

	Translation (cm)			Rotation (deg)			Focal length (pixel)	Image resolution
Scene	T_X	T_Y	T_Z	ω	ϕ	θ	f	
Building Hall	360	480	173	30, 35, 40, 45, 50	79, 83	3, 6, -3	1875	2688 × 1520
Courtyard	660	510	226	40, 45, 50, 55, 60	82, 86	3, 6, -3	1425	2048 × 1152
Volleyball Court	2223	1260	223	45, 50, 55, 60, 65	81, 84	3, 6, -3	1425	2048 × 1152

TABLE IV
ERRORS OF CAMERA PARAMETERS ESTIMATED WITH REAL IMAGES.

		Translation (cm)			Rotation (deg)			Focal length (pixel)
Scene	Method	ΔT_X	ΔT_Y	ΔT_Z	$\Delta\omega$	$\Delta\phi$	$\Delta\theta$	Δf
Building Hall	3D-calib+BA	19.98	29.73	12.68	1.58	1.12	0.66	133.98
	L-calib+BA	23.68	47.74	29.01	3.00	1.75	0.62	181.57
	5P-calib+BA	19.80	38.01	13.07	1.72	1.54	0.88	135.20
Courtyard	3D-calib+BA	27.65	23.74	11.61	2.44	2.75	0.57	79.95
	L-calib+BA	92.92	84.23	29.41	2.70	2.47	0.71	207.04
	5P-calib+BA	45.37	32.26	16.79	2.41	2.76	0.68	118.16
Volleyball Court	3D-calib+BA	10.14	6.18	7.32	2.30	2.68	0.24	22.81
	L-calib+BA	12.18	23.73	8.80	2.54	3.22	0.31	46.48
	5P-calib+BA	12.87	7.78	7.49	2.38	2.66	0.28	18.54

Comparing the results among three scenes, one can observe that higher errors are yielded in the courtyard scene than in the other scenes. This is because the reference object is relatively small in the courtyard images, as shown in Fig. 20, so that little errors of the reference point positons may lead to large deviation in the estimation results.

This experiment validates the applicability of our proposed method in different types of scenes, and the results confirm that 5P-calib+BA is able to estimate camera parameters close to those yielded by 3D-calib+BA. Based on the experiments with real images and simulated data in this and the previous sub-sections, we suggest some directions that may yield better estimation of camera parameters for a single image. (i) Use a reference object which is relatively large in the image. (ii) Select the reference points which can be located with high accuracy prior to those with inaccuracy. (iii) If sufficient reliable reference points can be located in the image, use 3D-calib with BA refinement to estimate camera parameters, since it tends to perform better. However, there exist the cases that 3D-calib may fail to deal with due to insufficient visible reference points, e.g., Figs. 11(b), 12(b), and 14(b). For such cases, 5P-calib can provide an alternative solution to obtain camera parameters with satisfying accuracy.

VIII. CONCLUSION

Camera calibration has been extensively researched in photogrammetry and, more recently, in computer vision. Various techniques using 3D reference objects (two or more orthogonal planes), 2D planes (planar patterns undergoing some motions), 1D sticks (points aligned on a line), or 0D features (self-calibration using scene feature points) are proposed in the literature. In this paper, we propose a novel geometric analysis-based method capable of directly estimating the position, orientation, and focal length of a zooming camera from just one image of five point correspondences. The proposed method is easy to implement, without requiring the complex linearization technology nor tedium computation of large matrixes. We also demonstrate how to apply the proposed method to sports videos, showing good practicability.

The proposed calibration method is examined and evaluated on both computer simulated data and real images, and is compared with a typical 3D object-based calibration and Miyagawa's approach [37], which also uses five point correspondences. Very encouraging results are obtained, demonstrating that our method can estimate camera parameters with lower errors than Miyagawa's approach. With bundle adjustment, our method can achieve good performance close to that of the 3D object-based method, even though only five point correspondences from a single image are used.

There are some limitations in the proposed calibration method that open the doors for further exploration. First, assuming a perspective camera model, the proposed calibration method does not take lens distortion into consideration. However, lens distortion is a topic of great importance when discussing about camera calibration. Therefore, one of our future research directions is to address the distortion problem in depth and endeavor to reduce the accuracy deterioration caused by lens distortion. Besides, the intrinsic parameters cannot be estimated by the proposed approach so far, and there are still degeneration configurations, e.g., when the optical axis is perpendicular to the scene plane, which cannot be dealt with by the current scheme. Thus, our future work also includes (i) deriving the intrinsic parameters and (ii) designing different geometric derivations for degeneration configurations.

REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2003.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2016.2595319, IEEE Transactions on Circuits and Systems for Video Technology

[2] K.-H. Lo and J.-H. Chuang, "Vanishing point-based line sampling for real-time people localization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1209–1223, July, 2013.

[3] H. Arai, I. Miyagawa, H. Koike, and M. Haseyama, "Estimating the number of people in video sequence via geometrical model," in *Proc. IEEE Int. Conf. Pattern Recognit.*, pp. 1–4, 2008.

[4] J. Han, D. Farin, and P.H.N. de With, "A mixed-reality system for broadcasting sports video to mobile devices," *IEEE Multimedia*, vol. 18, issue 2, pp. 72–84, Feb. 2011.

[5] J. Han, D. Farin, and P.H.N. de With, "A real-time augmented reality system for sports broadcast video enhancement," in *Proc. ACM Int. Conf. Multimedia*, pp. 337–340, 2007.

[6] X. Yu, N. Jiang, L.-F. Cheong, H.-W. Leong, and X. Yan, "Automatic camera calibration of broadcast tennis video with applications to 3D virtual content insertion and ball detection and tracking," *Comput. Vis. Image Und.*, vol. 113, issue 5, pp. 643–652, May 2009.

[7] G. Wang, H.-T. Tsui, Z. Hu, and F. Wu, "Camera calibration and 3d reconstruction from a single view based on scene constraints," *Image Vis. Comput.*, vol. 23, issue 3, pp. 311-323, Mar. 2005.

[8] R.-Y. Tsai, "An efficient and accurate camera calibration technique for 3-D machine vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 364–374, 1986.

[9] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. Pattern Anal. Mach Intell*, vol. 14, no. 10, pp. 965–980, Oct. 1992.

[10] J. Heikkilä, "Geometric camera calibration using circular control points," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1066–1077, Oct. 2000.

[11] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, 1993.

[12] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[13] P. F. Sturm and S. J. Maybank, "On plane-based camera calibration: a general algorithm, singularities, applications," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. 432–437, 1999.

[14] T. Ueshiba and F. Tomita, "Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, pp. 966–973, 2003.

[15] L. Lucchese, "Geometric calibration of digital cameras through multiview rectification," *Image Vis. Comput.*, vol. 23, no. 5, pp. 517–539, May 2005.

[16] Z. Zhang, "Camera calibration with one-dimensional objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 892–899, July 2004.

[17] F.-C. Wu, Z.-Y. Hu, and H.-J. Zhu, "Camera calibration with moving one-dimensional objects," *Pattern Recognit.*, vol. 38, no. 5, pp.755–765, May 2005.

[18] L. Wang, F.-C. Wu, and Z.-Y. Hu, "Multi-camera calibration with one-dimensional object under general motions," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1–7, 2007.

[19] Q.-T. Luong and O. D. Faugeras, "Self-calibration of a moving camera from point correspondences and fundamental matrices," *Int. J. Comput. Vis.*, vol. 22, no. 3, pp. 261–289, Mar. 1997.

[20] L. Wang and F. Duan, "Zhang's one-dimensional calibration revisited with the heteroscedastic error-in-variables model," in *Proc. IEEE Int. Conf. Image Process.*, pp. 857–860, 2011.

[21] P. Hammarstedt, P. Sturm, and A. Heyden, "Degenerate cases and closed-form solutions for camera calibration with one-dimensional objects," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1, pp.317–324, 2005.

[22] F. Qi, Q. Li, Y. Luo, and D. Hu, "Constraints on general motions for camera calibration with one-dimensional objects," *Pattern Recognit.*, vol. 40, no. 6, pp. 1785–1792, Jun. 2007.

[23] B. Triggs, "Autocalibration and the absolute quadric," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 609–614, 1997.

[24] E. E. Hemayed, "A survey of camera self-calibration," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, pp. 351–357, 2003.

[25] H. Ackermann and K. Kanatani, "Robust and efficient 3-D reconstruction by self-calibration," in *Proc. IAPR Conf. Mach. Vis. Appl.*, pp. 178–181, 2007.

[26] B. Caprile and V. Torre, "Using vanishing points for camera calibration," *Int. J. Comput. Vis.*, vol. 4, no. 2, pp. 127–139, Mar. 1990.

[27] R. Cipolla, T. Drummond, and D. Robertson, "Camera calibration from vanishing points in images of architectural scenes," in *Proc. British Mach. Vis. Conf.*, vol. 2, pp. 382–391, 1999.

[28] D. Liebowitz and A. Zisserman, "Metric rectification for perspective images of planes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 482–488, 1998.

[29] L.-L. Wang and W.-H. Tsai, "Camera calibration by vanishing lines for 3-d computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.13, no. 4, pp. 370–376, Apr. 1991.

[30] L. Grammatikopoulos, G. Karras, E. Petsa, and I. Kalisperakis, "A unified approach for automatic camera calibration from vanishing points," in *Proc. ISPRS Commiss. V Symp.: Image Eng. Vis. Metrol.*, pp. 1–6, 2006.

[31] M.-C. Hu, M.-H. Chang, J.-L. Wu, and L. Chi, "Robust camera calibration and player tracking in broadcast basketball video," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 266–279, Apr. 2011.

[32] J. Han, D. Farin, and P. H. N. de With, "Broadcast court-net sports video analysis using fast 3-D camera modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1628-1638, Nov. 2008.

[33] H.-T. Chen, M.-C. Tien, Y.-W. Chen, W.-J. Tsai, and S.-Y. Lee, "Physics-based ball tracking and 3D trajectory reconstruction with applications to shooting location estimation in basketball video," *J. Vis. Commun. Image Represent.*, vol. 20, issue 3, pp. 204–216, Apr. 2009.

[34] H.-T. Chen, W.-J. Tsai, S.-Y. Lee, and J.-Y. Yu, "Ball tracking and 3D trajectory approximation with applications to tactics analysis from single-camera volleyball sequences," *Multimedia Tools Appl.*, vol. 6, issue 3, pp. 641–667, Oct. 2012.

[35] N. Inamoto and H. Saito, "Free viewpoint video synthesis and presentation of sporting events for mixed reality entertainment," in *Proc. ACM SIGCHI Int. Conf. Adv. Comput. Entertainm. Technol.*, pp. 42–50, 2004.

[36] Hawk-Eye Coaching Systems. [Online]. Available: http://www.hawkeyeinnovations.co.uk/page/coaching-systems

[37] I. Miyagawa, H. Arai, and H. Koike, "Simple camera calibration from a single image using five points on two orthogonal 1-D objects," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1528–1538, Jun. 2010.

[38] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment—a modern synthesis," *Vis. Algorithms: Theory Pract*, pp. 298–375, 2000.

[39] B. Triggs, "Camera pose and calibration from 4 or 5 known 3D points," in *Proc. Int. Conf. Comput. Vis.*, pp. 278–284, 1999.

[40] M. Bujnak, Z. Kukelova, and T. Pajdla, "A general solution to the p4p problem for camera with unknown focal length," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1–8, 2008.

[41] M. Bujnak, Z. Kukelova, and T. Pajdla, "Making minimal solvers fast," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1506–1513, 2012.

[42] K. Josephson and M. Byröd, "Pose estimation with radial distortion and unknown focal length," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2419–2426, 2009.

[43] T. Sattler, C. Sweeney, and M. Pollefeys, "On sampling focal length values to solve the absolute pose problem," in *Proc. Europ. Conf. Comput. Vis.*, pp. 828–843, 2014.

[44] M. Bujnak, Z. Kukelova, and T. Pajdla. "Making minimal solvers fast," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1506–1513, 2012.

[45] Y. Zheng, S. Sugimoto, I. Sato, and M. Okutomi, "A general and simple method for camera pose and focal length determination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 430–437, 2014.

[46] Y. Liu, D. Liang, Q. Huang, and W. Gao, "Extracting 3D information from broadcast soccer video," *Image Vis. Comput.*, vol. 24, no. 10, pp. 1146–1162, Oct. 2006.

Hua-Tsung Chen (M'07) received the B.S., M.S., and Ph.D. degrees in Computer Science and Information Engineering from National Chiao Tung University, Hsinchu, Taiwan in 2001, 2003, and 2009, respectively.

He is currently an Assistant Research Fellow with Information and Communications Technology Lab, National Chiao Tung University, Hsinchu, Taiwan. His research interests include computer vision, video signal processing, content-based video indexing and retrieval, multimedia information system, and music signal processing.