Int. J. Human-Computer Studies (2000) **52,** 879–898
doi:10.1006/ijhc.1999.0365
Available online at http://www.idealibrary.com on IDEAL®

An efficient camera calibration method for vision-based head tracking

K. S. Park and C. J. Lim

Man-Machine Production Systems Laboratory, Department of Industrial Engineering, Korea Advanced Institute of Science and Technology, 373-1, Kusong-dong, Yusong-gu, Taejon, 303-701, Korea. e mail: kspark@convex.kaist.ac.kr

(Received 10 March 1999; accepted in revised form 29 November 1999)

The aim of this study is to develop and evaluate an efficient camera calibration method for vision-based head tracking. Tracking head movements is important in the design of an eye-controlled human/computer interface. A vision-based head tracking system is proposed to allow the user's head movements in the design of the eye-controlled human/computer interface. We propose an efficient camera calibration method to track the three-dimensional position and orientation of the user's head accurately. We also evaluate the performance of the proposed method and the influence of the configuration of calibration points on the performance. The experimental error analysis results showed that the proposed method can provide more accurate and stable pose (i.e. position and orientation) of the camera than the direct linear transformation method which has been used in camera calibration. The results for this study can be applied to the tracking of head movements related to the eye-controlled human/computer interface and the virtual reality technology.

KEYWORDS: vision-based head tracking; eye-controlled human/computer interface; camera calibration.

1. Introduction

As computers become more powerful, the critical bottleneck in their use is often in the interface to the user rather than the computing power. One of the goals in human computer interaction research is to increase the communication bandwidth between the user and the computer (Jacob, 1993). An additional mode of communication between the user and the computer would be useful for better interface.

Eye-controlled input devices provide a high-bandwidth source of additional input from the user to the computer. That is, the computer identifies the point on the display at which the user is looking and uses that information as a communication means. This alternative input device may help disabled users who can move their eyes much more effectively than they can operate any other computer input device (Hutchinson, White, Martin, Reichert & Frey, 1989; Lacourse & Hludik, 1990) described a prosthetic device called the eye-gaze-response interface computer aid (Erica), which admittedly has some limitations. The user has to maintain his or her head in a nearly stationary position

1071-5819/00/050879 + 20 $35.00/0

because lateral head movements greater than 2 in in either direction cause the eye image to leave the camera field and also because head movements greater than few inches toward or away from the camera put the eye image out of focus (Hutchinson *et al*., 1989).

The removal of constraint on user's head is of great importance in order to realise natural and flexible experimental environment. By integrating eye and head position tracking devices, Park and Lee (1996) developed an eye-controlled human/computer interface (EHCI) based on the line of sight (LOS) and an intentional blink to invoke commands. To track the head pose (position and orientation), they used Polhemus' FASTRAK using the magnetic field. A transmitter was positioned above the head, and a receiver was attached to the head band. FASTRAK computes the receiver's pose with respect to the transmitter.

There are various position tracking methods—mechanical, magnetic and optical methods to name a few. Mechanical position trackers have a low lag, are much less sensitive to their environment than magnetic position trackers, and tend to be affordable. However, they had a small working volume, and their sociability is poor because mechanical linkages create motion restrictions.

Magnetic position trackers are generally very flexible since they are small enough so that they can be attached to heads. However, they suffer from several side-effects. First, magnetic interface from devices such as radios or monitors can cause erroneous readings. Second, large objects made of ferrous metals can interfere with the electromagnetic field, causing inaccuracies (Meyer, Applewhite & Biocca, 1992).

Optical position trackers are able to work over a large area, but they need to maintain an LOS from the set of reference points to the camera. They have enjoyed success when used on helmet-mounted displays in aircraft cockpits (Meyer *et al*., 1992). Nakamura, Kobayashi, Taya and Ishigami (1991) first tried an optical method for head monitoring but the applications of the system are restricted to experiments involving mild head movements because they calculated the head pose using the image difference (i.e. the parallar instead of the camera model). Kang (1998) proposed a hands-free navigation system in virtual reality environments by tracking the head. In this approach, only scaled distances and not absolute distances can be extracted because the affine camera model was used.

We designed a vision-based head tracking system which is robust to electromagnetic interference and ferrometallic objects. We can extract three-dimensional (3D) position and orientation of the head using the system. A camera is attached to a user's head band and takes the front view containing eight 3D reference points (retro-reflecting markers) fixed at the computer monitor. The reference points are captured by an image processing board and are used to calculate the camera pose. Small-sized high-resolution camera devices that can be easily mounted on the user's head are readily available. The proposed vision-based head tracking system is illustrated in Figure 1. To track the head, it is required continuously to calculate the camera pose with respect to the set of reference points. For future progress in applying the proposed system, a simple, rapid and accurate camera calibration method which monitors the camera pose continuously is indispensable.

We propose an efficient camera calibration method for providing the accurate pose of the camera. We also assess the performance of the proposed camera calibration method

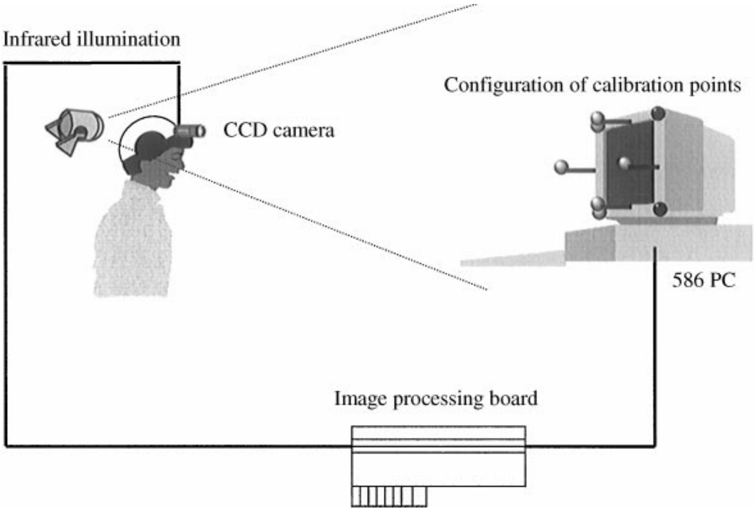

FIGURE 1. The schematic diagram of the vision-based head tracking system.

and the influence of the configuration of calibration points on the performance by real experiments.

This paper is organized as follows. In Section 2, the basic notions and literature reviews of camera calibration are presented. In Section 3, the details about the proposed camera calibration method are described. In Section 4, the relation between the accuracy of camera pose and the configuration of calibration points is discussed. In Section 5, the details about the experiments for validating the proposed camera calibration method are described. In Section 6, the experimental results are shown. Conclusions are drawn in Section 7.

2. Camera calibration

Camera calibration in the context of computer vision is the process of determining the geometric parameters of a mathematical camera model (Lenz & Tsai, 1988). In general, camera parameters can be divided into two categories, namely intrinsic parameters and extrinsic parameters. Intrinsic parameters are independent of the camera pose. They may include the effective focal length (the width and the height of photo-sensor cell), and the image centre (i.e. the image coordinates of the intersection of the optical axis and the image sensor plane). Extrinsic parameters are essentially the camera pose. Hence, they are independent of intrinsic parameters (Shih, Hung & Lin, 1996).

Usually, camera calibration is performed for two major purposes. One purpose is to identify the camera geometry of a 3D computer vision system. The other purpose is to calibrate a robot (either a robot arm or a robot head). Different vision tasks may demand the camera to be calibrated in different ways. If the camera rigidly attached to the user's head band observes the reference points on the computer monitor, the camera pose can be calculated from a camera calibration method and the pose of the user's head with the camera can be estimated. To track the head easily and precisely during video display terminal (VDT) work, an efficient camera calibration method is required.

Many techniques have been developed for camera calibration because of the strong demand of applications. Abdel-Aziz and Karara (1971) introduced a direct linear transformation (DLT) method, an approach that has been used in analytical photogrammetry. The DLT method that does not consider lens distortion is the one that estimates a DLT matrix, which consists of the composite parameters made by intrinsic and extrinsic camera parameters. If necessary, given the DLT matrix, camera parameters can be easily determined. Several researchers have reported on the acceptable accuracy afforded by the DLT method (Alem, Melvin & Holstein, 1978; Miller, Shapiro & McLaughlin, 1980). This method is used in biomechanical analysis to perform 3D space reconstruction from 2D film images.

Tsai (1987) proposed an efficient two-stage technique using the "radial alignment constraint" to consider lens distortion. Tsai's method involves a direct solution for most of the camera parameters and some iterative solutions for the remaining parameters. A drawback of Tsai's method has been mentioned by Weng, Cohen and Herniou (1992) and shown by Shih, Hung and Lin (1995). This method can be worse than the DLT method if lens distortion is relatively small. Hatze (1988) devised a modification of the conventional DLT method to increase its accuracy. In his assessment, the modified method worked far better than the conventional DLT method. Artificial neural nets (Wen & Schweitzer, 1991) and statistical methods (Czaplewski, 1992) have also been proposed to solve the camera calibration problem without specifying the camera model.

The criteria for measuring the effectiveness of a camera calibration method are autonomy, accuracy, simplicity, efficiency, flexibility and reliability (Ito & Ishii, 1994). To achieve these goals, it is extremely important to calculate extrinsic parameters with high accuracy under the condition of fixed intrinsic parameters because the intrinsic parameters calculated by the conventional DLT method tend to be changed although they should not when the camera moves. This means that the conventional DLT method can result in large fluctuations in the estimated camera parameters because intrinsic and extrinsic parameters are calculated simultaneously. In addition, we often come across unexpected discrepancies between the estimated camera parameters and the real ones, although the calculated calibration image coincides quite well with the real calibration image. Possibly, these errors can be attributed to ambiguity properties in the over determined system. We present a solution for suppressing these fluctuations and ambiguity properties in this paper. We propose an efficient calibration method based on the DLT method to provide the accurate pose of the camera.

3. Efficient camera calibration method

In Section 3.1, the camera model adopted in this paper is introduced. We describe a new camera calibration procedure based on the DLT method to provide accurate extrinsic parameters in Section 3.2. In Section 3.3, the importance of the image centre is described. In Section 3.4, the decomposition procedure of the composite camera parameters is described.

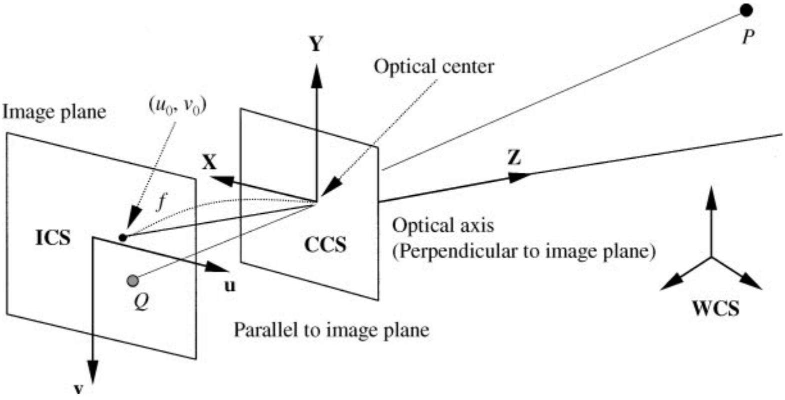

FIGURE 2. Pinhole camera model without lens distortion

3.1. CAMERA MODEL

Consider the pinhole camera model with the assumption that the camera performs perfect perspective projections, as shown in Figure 2. We do not consider lens distortion because the estimates of camera parameters are almost not affected if lens distortion is very small. All coordinate systems in this model are based on the right-handed coordinate system. In this model, the image plane is behind the optical centre like the real camera system and images are converted. The corresponding image point of P on the image plane would be Q (see Figure 2). The origin of the world coordinate system (WCS) is optional, the origin of the camera coordinate system (CCS) is the optical centre and the origin of the computer image coordinate system (ICS) is the centre of the frame memory coordinate [e.g. the origin of the ICS is set at (320, 240) for a 640×480 image].

3.1.1. Notations

- *Coordinates*

$\mathbf{r}_w = (x_w, y_w, z_w)$: coordinates of the 3D point P with respect to the WCS (in mm)
$\mathbf{r}_c = (x_c, y_c, z_c)$: coordinates of the 3D point P with respect to the CCS (in mm).
$\mathbf{s}_F = (u_F, v_F)$: 2D coordinates of Q with respect to the ICS (in mm).
$\mathbf{s}_I = (u_I, v_I)$: 2D coordinates of Q with respect to the ICS (in pixels) considering image centre.
(u_I, s, v_I, s): homogeneous coordinates of s_I:

- *Intrinsic parameters*

f: distance between the optical centre and the image plane, as shown in Figure 2, and referred to as “efficient focal length”.
(u_0, v_0): coordinates of the image centre with respect to the ICS (in pixels).
δ_u: horizontal pixel spacing (mm/pixel).
δ_v: vertical pixel spacing (mm/pixel).

- *Extrinsic parameters*

$$\mathbf{t}_C^W = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

: translation vector.

$$\mathbf{R}_C^W = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}$$

$$= \begin{bmatrix} \cos\phi_x \cos\phi_y & -\sin\phi_z \cos\phi_x \cos\phi_z \sin\phi_y \sin\phi_x & \sin\phi_z \sin\phi_x + \cos\phi_x \sin\phi_y \cos\phi_x \\ \sin\phi_z \cos\phi_y & \cos\phi_z \cos\phi_x + \sin\phi_z \sin\phi_y \sin\phi_x & -\cos\phi_z \sin\phi_x + \sin\phi_z \sin\phi_y \cos\phi_x \\ -\sin\phi_y & \cos\phi_y \sin\phi_x & \cos\phi_y \cos\phi_x \end{bmatrix}$$

: 3×3 rotation matrix determined by three Euler angles (ϕ_x, ϕ_y, ϕ_z).

- *Transformation matrices*

$$\mathbf{H}_C^F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{bmatrix}$$

: perspective projection from a 3D object point in the CCS to a 2D image point on the image plane.

$$\mathbf{T}_I^F = \begin{bmatrix} 1/\delta_u & 0 & u_0 \\ 0 & 1/\delta_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

: scaling and translation of 2D image coordinates.

$$\mathbf{T}_C^W = \begin{bmatrix} R_C^W & \mathbf{t}_C^W \\ 0 & 1 \end{bmatrix}$$

: translation and rotation from the WCS to the CCS.
The relationship between $\mathbf{r}_w$ and $\mathbf{s}_I$ can be expressed as a linear transformation equation

$$\tilde{\mathbf{s}}_I = \mathbf{H}\tilde{\mathbf{r}}_w \quad \text{i.e.} \quad \begin{bmatrix} u_I s \\ v_I s \\ s \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 \\ h_5 & h_6 & h_7 & h_8 \\ h_9 & h_{10} & h_{11} & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \tag{1}$$

where $\mathbf{H}$ is $\mathbf{T}_I^F \mathbf{H}_F^C \mathbf{T}_C^W$ and tilde (~) denotes homogeneous coordinates.

For convenience, let us define a **P** matrix, a **h** vector and a **q** vector as follows:

$$\mathbf{P} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_j & y_j & z_j & 1 & 0 & 0 & 0 & 0 & -u_j x_j & -u_j y_j & -u_j z_j \\ 0 & 0 & 0 & 0 & x_j & y_j & z_j & 1 & -v_j x_j & -v_j y_j & -v_j z_j \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{2N_{calib} \times 11},$$

$$\mathbf{h} = [h_1 \ h_2 \ h_3 \ h_4 \ h_5 \ h_6 \ h_7 \ h_8 \ h_9 \ h_{10} \ h_{11}]'_{11 \times 1},$$

$$\mathbf{q} = [\cdots \ u_j \ v_j \ \cdots]'_{16 \times 1}.$$

Then equation (1) can be combined into the following identification model:

$$\mathbf{Ph} = \mathbf{q}. \tag{2}$$

The 11 unknown composite camera parameters can be obtained using the least-squares method. Each data point pair $\{(x_i, y_i, z_i), (u_i, v_i)\}$ contributes two algebraic equations for the composite camera parameters. If there are more than six non-coplanar calibration points, we can obtain the 11 unknown composite parameters using the least-squares method.

3.2. CAMERA CALIBRATION PROCEDURE

In the conventional DLT method, due to the correlations between certain camera parameters, e.g. the correlation between the image centre and the camera orientation, the estimate of a set of camera parameters which minimizes a given criterion does not guarantee that physical camera parameter estimates are themselves accurate (Shih *et al.*, 1996). Kumar and Hanson (1990) and Lai (1993) showed that there was some linear dependency between the intrinsic parameters and the extrinsic parameters when the conventional DLT method is applied. Theoretically, intrinsic parameters should not be changed when the camera moves, but practically, these are not consistent because of the dependency. One should avoid directly estimating all the camera parameters simultaneously because of the correlations between the intrinsic parameters and the extrinsic parameters. Therefore, to improve the accuracy of extrinsic parameters, intrinsic parameters should be determined beforehand. This problem has not drawn much attention from computer vision because most computer vision applications require only accurate 3D measurements and do not care much about the values of physical parameters as long as their composite effect is satisfactory. However, in our application of vision-based head tracking, it is very important to choose a calibration method that provides the accurate pose of the camera.

We propose an efficient calibration method based on the DLT method to provide accurate extrinsic parameters. This method is to separate the camera calibration process into three independent steps. In the first step, we calibrate the image centre using the method of varying focal length. In the second step, we calibrate the effective focal length and the scale factor from the initial DLT matrix. In the third step, we calculate the

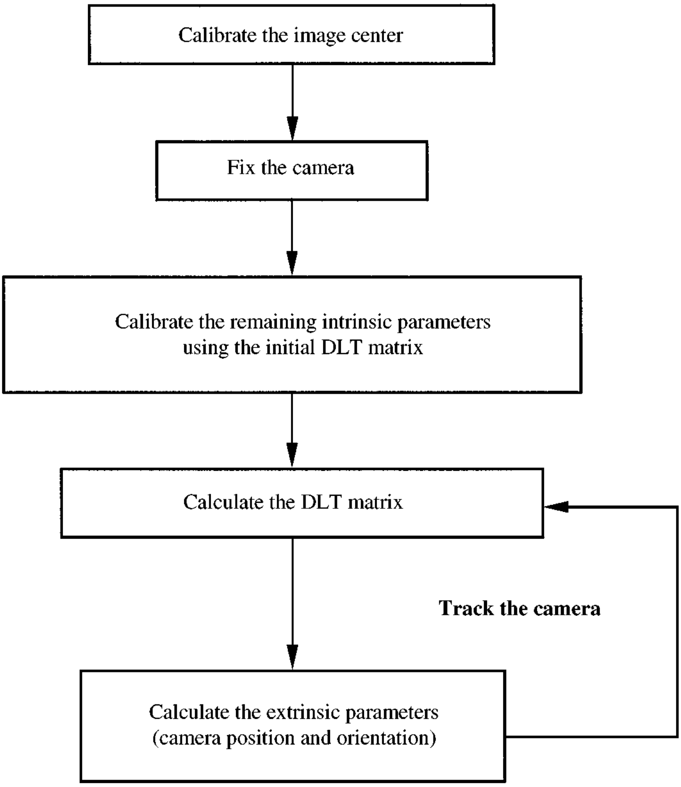
Calibrate the image center
Fix the camera
Calibrate the remaining intrinsic parameters using the initial DLT matrix
Calculate the DLT matrix
Track the camera
Calculate the extrinsic parameters (camera position and orientation)

FIGURE 3. Efficient camera calibration method.

camera pose from the DLT matrix using the calibrated intrinsic parameters. The procedure of the efficient camera calibration method is illustrated in Figure 3.

3.3. IMAGE CENTRE

The image centre is defined as the frame buffer coordinates (u_0, v_0) of the intersection of the optical axis with the image plane (Lenz & Tsai, 1988). It is often used as the origin of the imaging process and it appears in the perspective equations. It is a common practice in computer vision to choose the centre of the image frame buffer as the image canter. This is always fine for the analysis of 2D patterns. For 3D vision, the proper choice of the image centre can be critical to estimate the camera orientation (Lenz & Tsai 1988). However, the camera calibration methods that estimate the image centre together with the camera orientation suffer from the instability problem for the estimated parameters (Shih *et al.*, 1996). There are four different methods to determine the image centre independently of all the other camera parameters: a direct optical method, a method of varying focal length, a radial alignment method and a model fit method (Lenz & Tsai, 1988). We used a simple, inexpensive and accurate method for calibrating image centre, namely, the method of varying focal length, which is one of the methods frequently used in the field of computer vision.

3.4. DECOMPOSITION PROCEDURE

Given a set of 3D calibration points and their corresponding 2D image coordinates, the problem is to estimate the parameters of our camera model. Instead of estimating the parameters directly, we first estimate the composite parameters $h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9, h_{10}$ and h_{11}, then the composite parameters can be decomposed into the parameters of our camera model by the following method. Suppose that the estimated composite parameters are $h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9, h_{10}$ and h_{11}, we have the following 14 equation in 14 unknowns (the vertical pixel spacing δ_v is not included here because it is a known parameter when we use a solid-state camera): $r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8, r_9, t_1, t_2, t_3, \delta_u$ and f.

$$\begin{aligned}
r_1 &= (h_1 - h_9 u_0)\delta_u t_3/f, \\
r_2 &= (h_2 - h_{10} u_0)\delta_u t_3/f, \\
r_3 &= (h_3 - h_{11} u_0)\delta_u t_3/f, \\
t_1 &= (h_4 - u_0)\delta_u t_3/f, \\
r_4 &= (h_5 - h_9 v_0)\delta_v t_3/f, \\
r_5 &= (h_6 - h_{10} v_0)\delta_v t_3/f, \\
r_6 &= (h_7 - h_{11} v_0)\delta_v t_3/f, \\
t_2 &= (h_8 - v_0)\delta_v t_3/f, \\
r_7 &= t_3 h_9, \\
r_8 &= t_3 h_{10}, \\
r_9 &= t_3 h_{11}, \\
r_1^2 + r_4^2 + r_7^2 &= 1, \\
r_2^2 + r_5^2 + r_8^2 &= 1, \\
r_3^2 + r_6^2 + r_9^2 &= 1,
\end{aligned} \tag{3}$$

where the last three equations are the constraints of a rotation matrix. Substituting the first 11 equations into the last three, we have

$$\begin{bmatrix} (h_1 - h_9 u_0)^2 & (h_5 - h_9 v_0)^2 & h_9^2 \\ (h_2 - h_{10} u_0)^2 & (h_6 - h_{10} v_0)^2 & h_{10}^2 \\ (h_3 - h_{11} u_0)^2 & (h_7 - h_{11} v_0)^2 & h_{11}^2 \end{bmatrix} \begin{bmatrix} (\delta_u t_3/f)^2 \\ (\delta_u t_3/f)^2 \\ t_3^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{4}$$

By solving Equation (4), t_3, δ_u and f can be found, and we can easily obtain all the extrinsic parameters in Equations (3).

4. Error analysis

The accuracy of camera calibration is influenced by the camera model, the method of obtaining solutions, the 2D image measurement noise, and the configuration of calibration points (Zhang, Nomura & Fujii, 1995). The configuration concerns a number of factors, such as the number of calibration points, the distance between the camera and the calibration point, the distribution of calibration points, the calibrator depth, and so on. In this paper, we address the relation between the accuracy of camera parameters. We assess the influence of the configuration of calibration points on the performance of the proposed camera calibration method by real experiments.

4.1. THE NUMBER OF CALIBRATION POINTS

The error in estimating the 2D image coordinates of the calibration points is one of the sources of the camera calibration error. Shih *et al.* (1996) showed that the expectation of the average square 2D prediction error ε_n^2 is

$$\varepsilon_n^2 = \frac{11\sigma^2}{N_{calib}} \quad \text{(in pixels)},$$

where σ^2 is the variance of the 2D image measurement noise, and $N_{calib.}$ is the number of calibration points. The accuracy improves as the number of calibration points used in the configuration becomes larger, as shown in Figure 4, although the increase is very small. We selected eight calibration points because the average square 2D prediction error

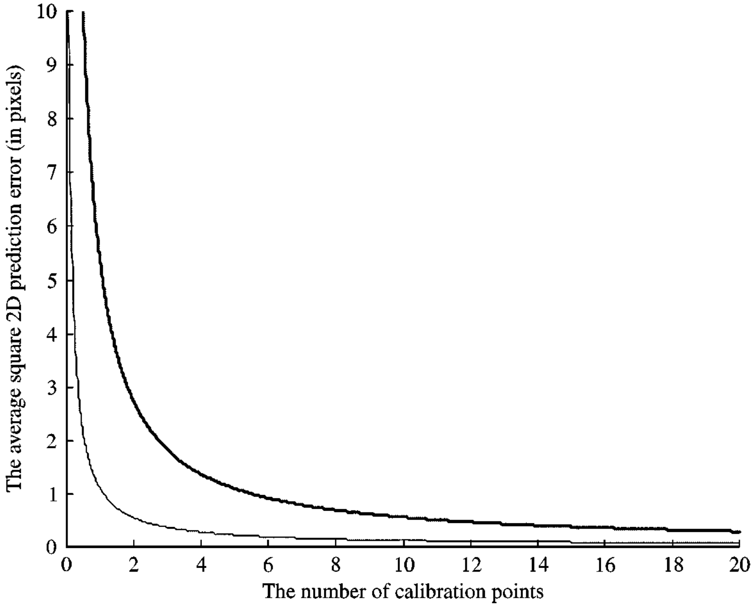

FIGURE 4. The average square 2D prediction error (in pixels) as a function of the number of calibration points.

under 1 pixel is sufficient for our application and the variance of the 2D image measurement noise in our vision system is less than 0.5.

4.2. THE DISTANCE BETWEEN THE CAMERA AND THE CALIBRATION POINT

From the perspective projection,

$$u_F = f\frac{x_c}{z_c}, \quad v_F = f\frac{y_c}{z_c}.$$

We can obtain the following results if we take the partial derivatives of u_F, v_F with respect to x_c, y_c, z_c:

$$\frac{\partial u_F}{\partial x_c} = \frac{\partial}{\partial x_c}\left(f\frac{x_c}{z_c}\right) = f\frac{1}{z_c}, \tag{5}$$

$$\frac{\partial u_F}{\partial y_c} = \frac{\partial}{\partial y_c}\left(f\frac{x_c}{z_c}\right) = 0,$$

$$\frac{\partial u_F}{\partial z_c} = \frac{\partial}{\partial z_c}\left(f\frac{x_c}{z_c}\right) = f\frac{x_c}{z_c^2}, \tag{6}$$

$$\frac{\partial v_F}{\partial x_c} = \frac{\partial}{\partial x_c}\left(f\frac{y_c}{z_c}\right) = 0,$$

$$\frac{\partial v_F}{\partial y_c} = \frac{\partial}{\partial y_c}\left(f\frac{y_c}{z_c}\right) = f\frac{1}{z_c}, \tag{7}$$

$$\frac{\partial v_F}{\partial x_c} = \frac{\partial}{\partial z_c}\left(f\frac{y_c}{z_c}\right) = f\frac{y_c}{z_c^2}. \tag{8}$$

From Equations (5)–(8), we can notice that $\partial u_F/\partial x_c$, $\partial u_F/\partial z_c$, $\partial v_F/\partial y_c$ and $\partial v_F/\partial z_c$ decrease as z_c is increased. If the distance between the camera and the calibration point is large, though the position of the calibration point is changed considerably in the CCS, the position in the ICS is not changed much. Conversely speaking, the 2D image measurement noise in the ICS greatly affects the 3D coordinates that are transformed into the CCS. Therefore, the estimation error of camera parameters is proportional to the average of the distances between the camera and the calibration points.

4.3. THE STABILITY OF THE CAMERA PARAMETERS ESTIMATION (THE DISTRIBUTION OF CALIBRATION POINTS)

The overdetermined linear system [Equation (2)] must be solved repeatedly to estimate the camera pose continuously. In Equation (2), $\mathbf{P}$ is a 16×11 matrix, $\mathbf{h}$ is an 11×1 vector, $\mathbf{q}$ is a 16×1 vector and $\mathbf{P}_{16 \times 11}$ and $\mathbf{q}_{11 \times 1}$ can be calculated from the calibration point data according to the camera model. The least-squares method is used for a wide variety of applications. A serious problem that may dramatically impact the usefulness of this method is multicollinearity. Multicollinearity implies near-linear dependency among the coefficients' vectors (Montgomery & Peck, 1992). The coefficients' vectors in our linear system are the columns of the $\mathbf{P}$ matrix, so clearly an exact linear dependency would result in a singular $\mathbf{P'P}$. The presence of near-linear dependencies can dramatically impact the ability to estimate the $\mathbf{h}$ vector. Suppose we use the unit length scaling so that the $\mathbf{P'P}$ matrix will be in the form of a correlation matrix $\mathbf{W'W}$, $Var(\hat{\mathbf{h}})$ can be written as $(\mathbf{W'W})^{-1}\sigma^2$ (Montgomery & Peck, 1992). The least-squares method, when strong multicollinearity is present, generates poor estimates, and the estimated values of the $\mathbf{h}$ vector are often very sensitive to the data in the particular sample collected. The diagnosis and treatment of multicollinearity is one of the important aspects of parameter estimation using the least-squares method.

For detecting multicollinearity, there are several techniques such as an examination of the correlation matrix, variance inflation factors and an eigensystem analysis of the correlation matrix. We adapted the eigensystem analysis of the correlation matrix to detect multicollinearity. The normal equation of Equation (2) is $\mathbf{P'Ph} = \mathbf{P'q}$. Let $\mathbf{P'P}$ be $\mathbf{A}_{11\times 11}$ and $\mathbf{P'q}$ be $\mathbf{b}_{11\times 1}$, then due to the 2D image measurement noise, $\mathbf{A}$ becomes $\mathbf{A}+\delta\mathbf{A}$ and $\mathbf{b}$ becomes $\mathbf{b}+\delta\mathbf{b}$, consequently, $\mathbf{h}$ becomes $\mathbf{h}+\delta\mathbf{h}$. From $(\mathbf{A}+\delta\mathbf{h})(\mathbf{h}+\delta\mathbf{h}) = \mathbf{b}+\delta\mathbf{b}$, we have $\mathbf{Ah}+\mathbf{A}\delta\mathbf{h}+\delta\mathbf{Ah}+\delta\mathbf{A}\delta\mathbf{h} = \mathbf{b}+\delta\mathbf{b}$. Then, we have the following results: $(\mathbf{A}+\delta\mathbf{A})\delta\mathbf{h} = \delta\mathbf{b}-\delta\mathbf{Ah}$ and $\delta\mathbf{h} = (\mathbf{A}+\delta\mathbf{A})^{-1}(\delta\mathbf{b}-\delta\mathbf{Ah})$. Let $\|\cdot\|$ denote the norm operator, if $\|(\delta\mathbf{A})\mathbf{A}^{-1}\| < 1$, then according to the condition of equation theorem (Nobel & Daniel, 1977),

$$\frac{\|\delta\mathbf{h}\|}{\|\mathbf{h}\|} = \frac{1}{1-\|(\delta\mathbf{A})\mathbf{A}^{-1}\|}\,Cond(\mathbf{A})\left\{\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}+\frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|}\right\},$$

where $Cond(\mathbf{A}) = \|\mathbf{A}\|\,\|\mathbf{A}^{-1}\|$is called the condition number of $\mathbf{A}$. It is obvious that the greater the $Cond(\mathbf{A})$, the larger the coefficients perturbation of $\mathbf{A}$ and $\mathbf{b}$ may cause the error of the solution $\mathbf{h}$. When the condition number is small, the overdetermined linear system is said to be well-conditioned, whereas when the condition number is large, the overdetermined linear system is said to be ill-conditioned. We can avoid multicollinearly by arranging the calibration points. We should select the configuration of calibration points with smaller condition number to obtain more stable results. Position data that are close together in the 3D space are often similar in the 2D image plane than those that are far apart. Therefore, the calibration points should be uniformly distributed in the 3D space and also in the 2D image plane.

4.4. THE CALIBRATOR DEPTH

Let us define the calibrator depth as the distance between the nearest calibration point and the farthest calibration point to the camera. The 2D image coordinates of the 3D calibration points with larger calibrator depth are more sensitive to the camera movement. Theoretically, we can calculate more accurate pose of the camera as we set the calibrator depth larger because the change rate of the calibration point images in the 2D image plane with respect to the camera movement is larger. Practically, the calibration points out of the depth of field [the distance between the nearest point and the farthest point from the camera which are imaged with acceptable sharpness (Hallert, 1960)] form blurred or defocused images. The blurring effect is the source of errors in calculating the 2D image coordinate of the 3D calibration point (the centre of hemispherical form marker). Therefore, one should select a suitable calibrator depth that comprises the "sensitive" effect and the "blurring" effect.

5. Experiments

The following series of experiments were performed to assess the accuracy and the stability of the proposed camera calibration method, and the influence of the configuration of calibration points. The experimental set-up used in our experiments is illustrated in Figure 5. It consists of a configuration of calibration points on a computer monitor, a CCD camera (SONY XC-77) with an 8 mm focal length lens attached an infrared pass filter, an image processing board (Samsung MVB03), a 586 PC as a processing unit, an infrared illumination and a 3D stage controller. We used the 3D stage controller which has higher accuracy (resolution: 0.01 mm) than our vision system to regard the controlled pose as the known pose. The CCD camera takes the images of the calibration points. The effective part of the CCD sensor array in the camera has 640 × 480 pixels, and the image processing board (digitizer) gives digital image with 8 bytes/pixel. The coordinates of the calibration points in the ICS are calculated by a digital signal processing program (DSP), and the data are transferred from the image processing board to the computer through AT bus. The coordinates of the calibration points on the captured image can be extracted easily since the markers have prominent brightness. We use the centre of region of the marker image as the coordinate of the calibration point. The computer calculates the camera poses using camera calibration methods. The accuracy and the stability of the proposed method compared with those of the conventional DLT method.

To assess the influence of the configuration of calibration points, different configurations were investigated. We designed four configurations of calibration points according Table 1. There are two factors with two levels. One is uniformity and the other is the calibrator depth as shown in Table 1. The designed configurations of calibration points are shown in Figure 6. Each configuration of calibration points consists of eight points

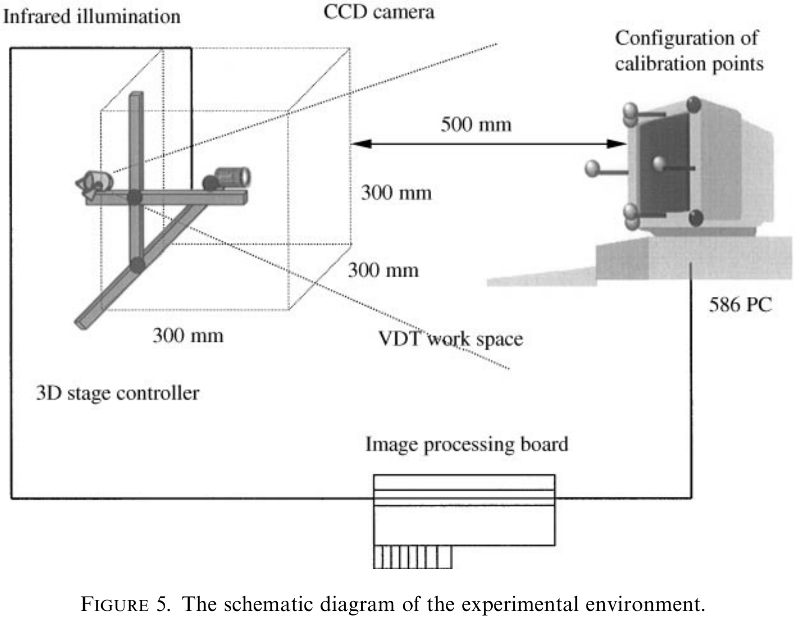

FIGURE 5. The schematic diagram of the experimental environment.

TABLE 1
The design of the configuration of calibration points

Uniformity	Calibration depth (mm)	
	100	200
Weak	Configuration #1	Configuration #3
Strong	Configuration #2	Configuration #4

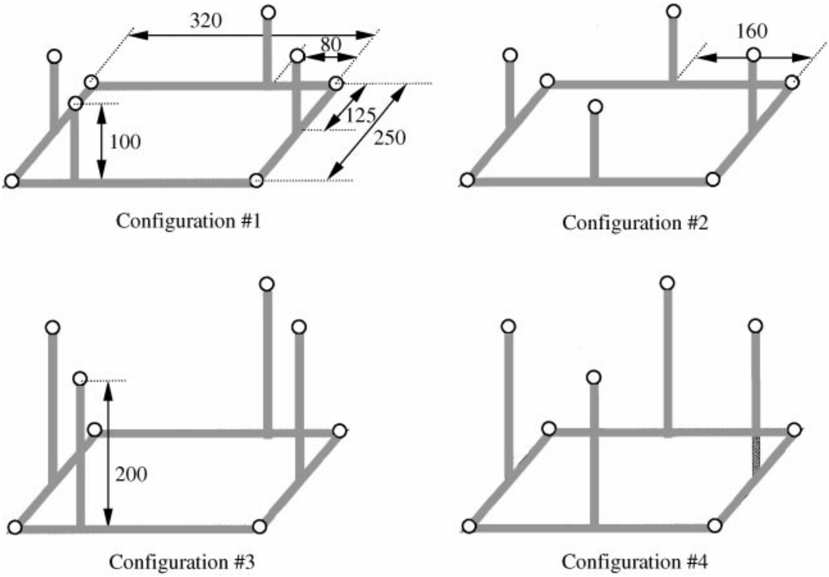

FIGURE 6. The schematic diagram of the designed configurations of calibration points (unit: mm).

and the position of each point is illustrated in Figure 6. Each of the calibration points is the centre of hemispherical form markers with retro-reflective property and a 5 mm radius. For each configuration of calibration points, the accuracy test and the stability test were performed.

5.1. ACCURACY TEST

To assess the accuracy, it is important that the known poses of the camera are independent. Therefore 30 poses were selected at random in the space, as shown in Figure 5, where the user's head is frequently located during VDT work. The camera was controlled to the selected poses. At each pose, we estimated the camera pose first using the conventional DLT method, and then using the proposed method. The estimated poses are compared with the known poses, and the accuracy was computed in terms of the degree to which the estimated poses agree with the true poses.

5.2. STABILITY TEST

To assess the stability, one of the 30 poses was selected and the camera was positioned to the selected pose. We estimated the camera pose 30 times first using the conventional DLT method, and then using the proposed method. The stability was computed in terms of the degree to which the estimated pose for the camera agrees with the corresponding mean estimated pose for the camera. The condition number, as described in Section 4.3, was calculated to identify the relation to the stability.

6. Experimental results

To evaluate the accuracy and the stability of camera calibration methods, the 3D position error and the 3D orientation error defined below are used as error measures. The 3D position error is defined as the Euclidean distance between the known 3D position and the estimated 3D position of the camera. The 3D orientation error is defined as the angle between the known orientation vector and the estimated orientation vector of the camera. The error measures used in this paper are illustrated in Figure 7. We also use the normalized 3D position error and the normalized 3D orientation error that are normalized by the average of the distances between the camera and the calibration points because the 3D position error and the 3D orientation error tend to increase as the distance is increased.

In Table 2 and Figure 8, the normalized 3D position error and the normalized 3D orientation error are summarized. The errors increased in the order configuration #4, configuration #3, configuration #2, configuration #1. This result is presumably due to the effect of the calibrator depth and the uniformity. From these results, we can recommend that configuration #4 will provide the most accurate and stable extrinsic parameters among the four configurations.

In Table 3, the condition number increased in the order configuration #4, configuration #2, configuration #3, configuration #1. This implies that configuration #4 may provide the most stable extrinsic parameters among the four configurations.

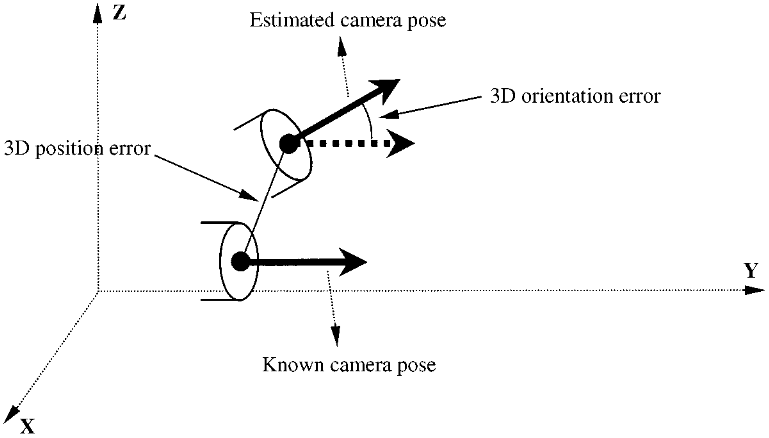

FIGURE 7. The error measures used in this paper.

TABLE 2
The accuracy of the proposed method and the DLT method

Error measure	Method	Configuration #			
		4	3	2	1
Average of normalized 3D position error (deg)	Proposed DLT	4.209E − 01 1.142E + 00	6.461E − 01 2.041E + 00	9.530E − 01 3.529E + 00	1.130E + 00 4.170E + 00
S.D. of normalized 3D position error	Proposed DLT	1.202E − 01 3.520E − 01	2.027E − 01 5.120E − 01	2.728E − 01 8.520E − 01	4.088E − 01 1.424E + 00
Average of normalized 3D orientation error (deg/mm)	Proposed DLT	3.801E − 04 1.517E − 03	5.749E − 04 1.971E − 03	6.928E − 04 2.204E − 03	7.197E − 04 2.812E − 03
S.D. of normalized 3D orientation error	Proposed DLT	1.340E − 04 2.432E − 04	2.989E − 04 5.102E − 04	3.285E − 04 8.430E − 04	4.660E − 04 9.291E − 04

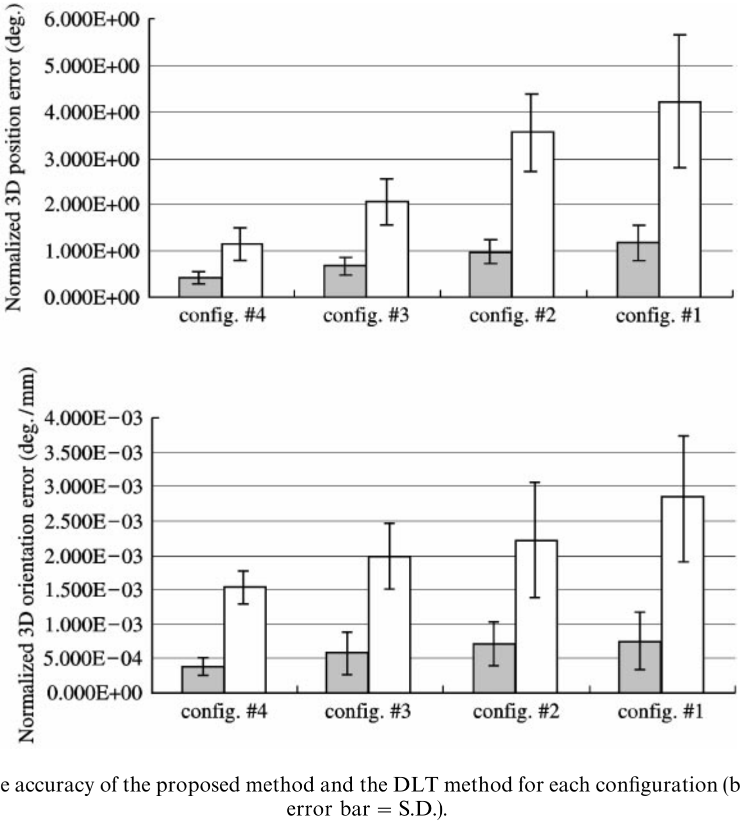

FIGURE 8. The accuracy of the proposed method and the DLT method for each configuration (bar = average, error bar = S.D.).

TABLE 3
The condition number for each configuration

Configuration #	4	3	2	1
Condition number	15.459	22.560	68.562	105.253

TABLE 4
The stability of the proposed method and the DLT method for each configuration

Error measure	Method	Configuration #			
		4	2	3	1
S.D. of 3D position	Proposed DLT	3.435E − 01 9.540E − 01	8.933E − 01 1.584E + 00	1.046E − 00 2.594E + 00	2.233E + 00 4.415E + 00
S.D. of 3D orientation	Proposed DLT	2.818E − 02 5.354E − 02	8.310E − 02 1.930E − 01	1.620E − 01 2.275E − 01	2.322E − 01 4.123E − 01

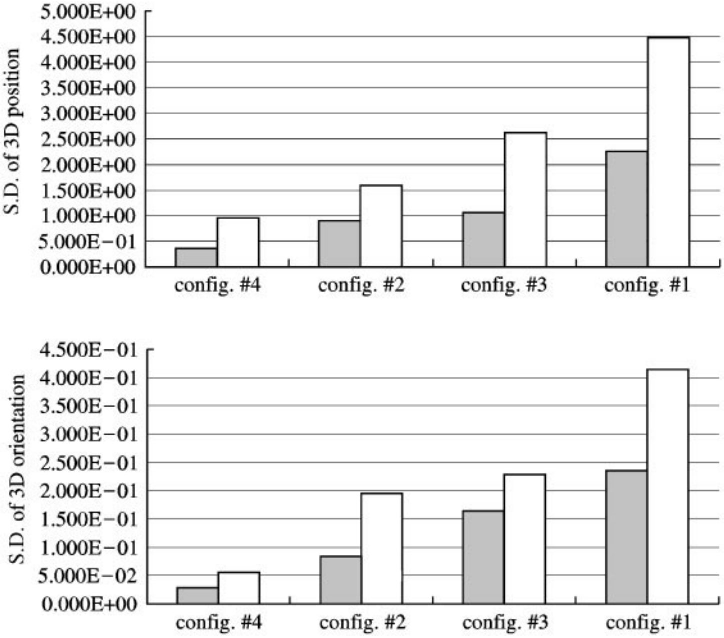

FIGURE 9. The stability of the proposed method and the DLT method for each configuration.

In Table 4 and Figure 9, the stability for each configuration is summarized. The standard deviation of the pose increased in the order configuration #4, configuration #2, configuration #3, and configuration #1. This result is consistent with the condition number, as described in Section 4.3.

From Figures 8 and 9, we can conclude that the proposed camera calibration method is more accurate and stable than the conventional DLT method when it is applied to the proposed head tracking system. When configuration #4 was adopted, the average of the normalized 3D position errors was about 4.209×10^{-1} degree, and the average of the normalized 3D orientation errors was about 3.801×10^{-4} degree/mm. This means that the proposed system will make 4.3 mm of the 3D position error and 0.25° of the 3D orientation error, on average, when the configuration of calibration points is 650 mm away from the camera. When 586 PC was used as a processing unit, the data rate was about 17 Hz.

7. Conclusions

This study was undertaken to design a vision-based head tracking system for the eye-controlled human/computer interface (EHCI). One objective was to develop an efficient camera calibration method for providing the accurate pose (position and orientation) of the camera. The other objective was to assess the performance of the proposed camera calibration method and the influence of the configuration of calibration points on the performance by real experiments. The experimental error analysis results showed that the proposed system makes 4.3 mm of the 3D position error and 0.25° of the 3D orientation error, on average, when the camera is 650 mm away from the configuration of calibration points. The results also showed that the proposed camera calibration method is far superior to the conventional DLT method when applied to the proposed vision-based head tracking system.

The configuration of calibration points is important to achieve accurate and stable estimation results. In real experiments, configuration #4 provides the most accurate and stable estimation results among the four configurations. Although the results reported here are for the calibration structure of specified dimensions, the recommendation has implications for studies attemping to calibrate the spaces of different dimensions. One should follow the guidelines listed below to achieve more accurate and stable estimation results.

- The number of calibration points should be selected as large as possible.
- The distance between the camera and the calibration point should be made as small as possible because the estimation error of camera parameters is proportional to the distance.
- The 3D and 2D coordinates of the calibration points should be uniformly distributed in the 3D space and also in the 2D image place.
- The calibrator depth should be made as large as possible within the depth of field.

However, to reduce the calibrator depth for practical application with as little errors as possible, we plan to use a multi-focal plane in our future research.

More recently, position trackers have been used to control computer-generated images in virtual reality applications. The user interacts with the virtual reality system through body movements; by moving the head, the user controls a computer-generated world. Since the proposed vision-based head tracking system can detect the head movements continuously, the results of this study show the possibility of application to the integral part of a virtual reality system.

References

ABDEL-AZIZ, Y. I. & KARARA, H. M. (1971) Direct linear transformation from comparator coordinates into object-space coordinates. In *Proceedings of ASP/UI Symposium on Close-Range Photogrammetry*, pp. 1–8. Urbana Champaign, IL: American Society of Photogrammetry.

ALEM, N. M., MELVIN, J. W. & HOLSTEIN, G. L. (1978). Biomechanics applications of Direct Linear Transformation in close-range photogrammetry. In *Proceedings of the 6th New England Bio-Engineering Conference*, pp. 202–206. Kingston, RI: Pergamon Press.

CZAPLEWSKI, R. L. (1992). Misclassification bias in areal estimates. *Photogrammetric Engineering and Remote Sensing*, **58,** 189–192.

HALLERT, B. (1960). *Photogrammetry*. New York: McGraw-Hill.

HATZE, H. (1988). High precision three-dimensional photogrammetric calibration and object space reconstruction using a modified DLT-approach, *Journal of Biomechanics*, **21,** 533–538.

HUTCHINSON, T. E., WHITE, K. P., MARTIN, W. N., REICHERT, K. C. & FREY, L. A. (1989). Human–computer Interaction eye-gaze input. *IEEE Transactions on System Man Cybernetics*, **19,** 1527–1534.

ITO, M. & ISHII, A. (1984). A non-iterative procedure for rapid and precise camera calibration. *Pattern Recognition*, **27,** 301–310.

JACOB, R. J. K. (1993). Eye-gaze computer interactions: what you look at is what you get. *IEEE Computer*, **26,** 65–67.

KANG, S. B. (1998). Hands-free navigation in VR environments by tracking the head. *International Journal of Human-Computer Studies*, **48,** 247–266.

KUMAR, R. & HANSON, A. R. (1990). Sensitivity of the pose refinement problem to accurate estimation of camera parameters. In *Proceedings of International Conference on Computer Vision* (*ICCV'90*), pp. 365–369.

LACOURSE, J. R. & HLUDIK, F. C. (1990). An eye movement communication-control system for the disabled. *IEEE Transactions on Biomedical Engineering*, **37,** 1215–1220.

LAI, J. Z. C. (1993). On the sensitivity of camera calibration. *Image Vision and Computing*, **11,** 656–664.

LENZ, R. K. & TSAI, R. Y. (1988). Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10,** 713–720.

MEYER, K., APPLEWHITE, H. L. & BIOCCA, F. A. (1992). A survey of position trackers. *Presence*, **1,** 173–200.

MILLER, N. R., SHAPIRO, R. & MCLAUGHLIN, T. M. (1980). A technique for obtaining spatial kinematic parameters of biomechanical systems from cinematographic data. *Journal of Biomechanics*, **13,** 535–547.

MONTGOMERY, D. C. & PECK, E. A. (1992). Multicollinearity. *Introduction to Linear Regression Analysis*, pp. 305–365. New York: John Wiley & Sons.

NAKAMURA, H., KOBAYASHI, H., TAYA, K. & ISHIGAMI, S. (1991). A design of eye movement monitoring system for practical environment. In *Proceedings of SPIE*, vol. 1456, pp. 226–238, San Jose, CA: The Society of Photo-Optical Instruction Engineers.

NOBEL, B. & DANIEL, J. W. (1977) *Application Linear Algebra*, London: Prentice Hall, Inc.

PARK, K. S. & LEE, K. T. (1996). Eye-controlled human/computer interface using the line-of-sight and the intentional blink. *Computers & Industrial Engineering*, **30,** 463–473.

SHIH, S. W., HUNG, Y. P. & LIN, W. S. (1995). When should we consider lens distortion in camera calibration. *Pattern Recognition*, **28,** 447–461.

SHIH, S. W., HUNG, Y. P. & LIN, W. S. (1996). *Accuracy analysis on the estimation of camera parameters for active vision systems.* Technical Report, TR-IIS-96-003, Institute of Information & Science, Academia Sinica, Nankang, Taipei, Taiwan.

TSAI, R. T. (1987). A vesatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, **RA-3,** 323–344.

WEN, J. & SCHWEITZER, G. (1991). Hybrid calibration of CCD camera using artificial neural nets. In *Proceedings of IEEE International Joint Conference on Neural Networks*, vol. 1, pp. 337–342.

WENG, J., COHEN, P. & HERNIOU, M. (1992). Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, **14,** 965–980.

ZHANG, D., NOMURA, Y. & FUJII, S. (1995). Error analysis and optical setup on camera calibration. In *Proceedings of Asian Conference on Computer Vision* (*ACCV'95*), pp. 210–214. Singapore.

Paper accepted for publication by Associate Editor, Professor P. Barker