A New Active Visual System for Humanoid Robots

De Xu, *Member, IEEE*, You Fu Li, *Senior Member, IEEE*, Min Tan, and Yang Shen

***Abstract*—In this paper, a new active visual system is developed, which is based on bionic vision and is insensitive to the property of the cameras. The system consists of a mechanical platform and two cameras. The mechanical platform has two degrees of freedom of motion in pitch and yaw, which is equivalent to the neck of a humanoid robot. The cameras are mounted on the platform. The directions of the optical axes of the two cameras can be simultaneously adjusted in opposite directions. With these motions, the object's images can be located at the centers of the image planes of the two cameras. The object's position is determined with the geometry information of the visual system. A more general model for active visual positioning using two cameras without a neck is also investigated. The position of an object can be computed via the active motions. The presented model is less sensitive to the intrinsic parameters of cameras, which promises more flexibility in many applications such as visual tracking with changeable focusing. Experimental results verify the effectiveness of the proposed methods.**

***Index Terms*—Active vision, bionic vision, humanoid robot, positioning, visual system.**

I. Introduction

THE PINHOLE model for cameras has been widely used in robot visual systems [1]. Generally, the parameters in the camera model need to be calibrated to perform visual measurement or control. The inherent parameters of a camera, such as the focus length, the principal point, and the magnification coefficients from the imaging plane coordinates to the image coordinates, are referred to as intrinsic parameters. The external parameters such as the relative positions and orientations of cameras are the extrinsic parameters. In many applications such as visual positioning [2], [3] and motion estimation [4], only the intrinsic parameters are of concern. On the other hand, the intrinsic and extrinsic parameters are important in applications with stereovision [5]. Up to now, the calibration for intrinsic parameters of a camera [5] has been well studied including the use of a special planar pattern [6], [7]. Although the methods are effective, their calibrating process is, in general, tedious and prone to errors.

Manuscript received March 5, 2007; revised July 5, 2007. This work was supported in part by the National Natural Science Foundation of China under Grant 60672039, the Research Grants Council of Hong Kong under Project CityU117605, and the National Key Fundamental Research and Development Project of China (973, No. 2002CB312200). This paper was recommended by Associate Editor H. Zhang.

D. Xu, M. Tan, and Y. Shen are with The Key Laboratory of Complex System and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China (e-mail: sdxude@yahoo.com; tan@compsys.ia.ac.cn; yang.shen@mail.ia.ac.cn).

Y. F. Li is with the Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong (e-mail: meyfli@cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2007.912082

To reduce the influence of the errors in camera calibration on visual control, some researchers developed the image-based visual servoing (IBVS) [1], [8] and hybrid visual servoing methods [9]. The camera's parameters are not separately estimated in IBVS, but included in the estimation of the image Jacobian matrix. With the camera parameters in the feedback loop of the image features, the influence of errors in camera calibration is reduced, but still exists.

Self-calibrating methods have been studied to eliminate the need for special patterns and to increase the adaptability of the visual system. One category of such calibration is based on special motions of the camera [10]. Another is based on the environment information such as parallel lines [11]–[13]. Recently, attention has focused on uncalibrated visual servoing (UCVS). In fact, the cameras in some UCVS systems are self-calibrated [14]. The methods in some UCVS systems belong to IBVS since cameras' parameters are not individually estimated, but combined into the estimation of the image Jacobian matrix [15]. Some researchers pursue the visual control without camera parameters [16]–[18]. For instance, Shen *et al.* [16] limited the workspace of the end-effector on a plane that is vertical to the optical axis of the camera to eliminate the camera parameters in the image Jacobian matrix. A visual control method based on the epipolar line and the cross ratio invariance was developed with two uncalibrated cameras in [18]. It did not use camera parameters, and the working space of the end-effector was in 3-D Cartesian space. However, this method was limited to approaching task.

The results of traditional visual measurements are dependent much on cameras' parameters, particularly the intrinsic parameters. In general, the focus of a camera is fixed, which heavily limits its flexibility in practical applications such as visual tracking. In addition, a camera needs to be calibrated before it is to be used for a new task. Obviously, the visual measurement and control methods that are insensitive to camera intrinsic parameters would be much more flexible and convenient to use than traditional ones.

The motivation of this paper is to develop a new visual system that is insensitive to the property of the cameras. An active visual system as well as its positioning method is designed to conduct visual measurement in the center areas of the cameras, which is insensitive to the intrinsic parameters. With the geometry information of our visual system, the position of an object can be determined even if the intrinsic parameters of the cameras are not available. The rest of this paper is organized as follows. The bionic visual models are introduced in Section II. One model is for the humanoid robot with a head of two degrees of freedom (DOFs). Another is a general model for any mobile robots. In Section III, the relative positioning for multiple objects is discussed. Section IV investigates the errors

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.

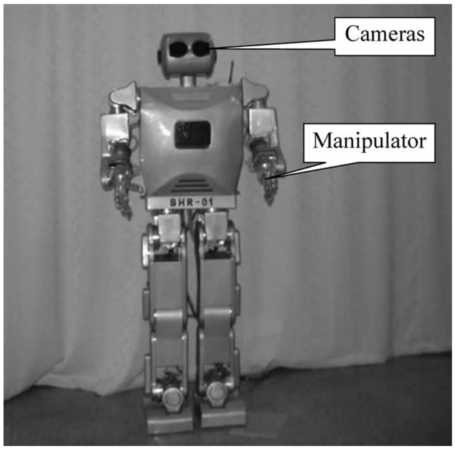


Fig. 1. Structure of a humanoid robot.

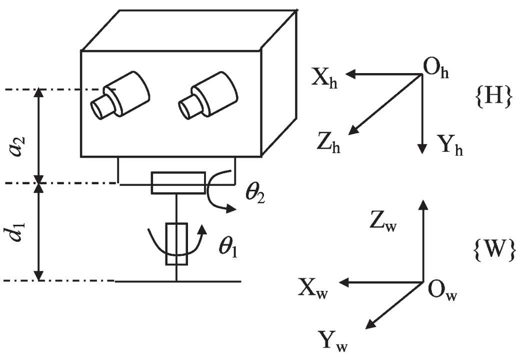


Fig. 2. Sketch of the neck and the head.

for the two proposed models. The calibration method for the initial directions of the optical axes of the cameras is provided in Section V. The experimental results are given in Section VI. Finally, this paper is concluded in Section VII.

II. Bionic Visual Model

A. Visual System for a Humanoid Robot

A humanoid robot has a typical configuration of the visual system as follows [19]. There are two cameras mounted on the head of the robot, which serve as the eyes. An eye-to-hand system is formed with these two cameras and a manipulator. The head has two DOFs: yaw and pitch [20]. The cameras and the head can be taken as an eye-in-hand system. With the two DOFs, the head can work as an active vision system (Fig. 1).

The sketch of the neck and the head of a humanoid robot is given in Fig. 2. The first joint is responsible for yawing, and the second one for pitching. The world frame W for the head is assigned at the connect point of the neck and the body. The head frame H is assigned at the midpoint of the two cameras.

B. Bionic Visual Model for a Humanoid Robot

The two cameras can simultaneously yaw in opposite directions to stare at an object. In the initial state of the two cameras, they are well mounted so that their optical axes are almost parallel. Therefore, the line connecting the two cameras is on the plane formed by the two optical axes. The following

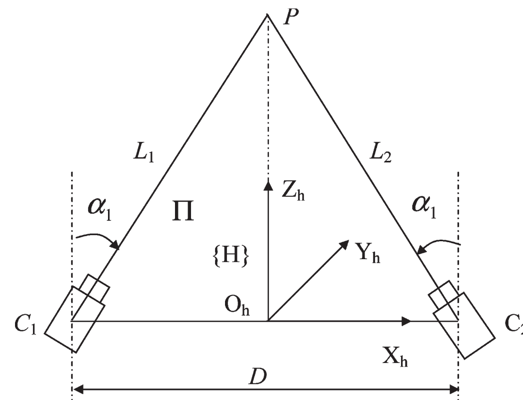


Fig. 3. Principle of visual positioning.

symbols are defined to describe the cameras (see also Fig. 3). L_1 denotes the optical axis of a camera C_{a1}. C_1 is its optical principal point. L_2 and C_2 indicate the optical axis and the optical principal point, respectively, of another camera C_{a2}. Π denotes the plane formed by L_1 and L_2. The position of a point P is expressed as $[x_h, y_h, z_h]$ in frame H, and $[x_w, y_w, z_w]$ in frame W.

For a point P, it can be adjusted to be on the plane Π with the change in θ_2. Then, it can be on the perpendicular bisector of line C_1C_2 on the plane Π with the adjustment of θ_1. With simultaneous yawing in opposite directions for the two cameras, the images of point P can be placed at the center positions of the image planes of the two cameras.

The transformation matrix from frame W to H is given in (1) according to the Denavit–Hartenberg (D-H) parameters model, where d_1 and a_2 are the D-H parameters of the neck's joints. θ_1 and θ_2 are the joint angles of the two joints.

$$ {}^{w}T_h = \begin{bmatrix} \cos\theta_1 & -\sin\theta_1\sin\theta_2 & -\sin\theta_1\cos\theta_2 & a_2\sin\theta_1\sin\theta_2 \\ \sin\theta_1 & \cos\theta_1\sin\theta_2 & \cos\theta_1\cos\theta_2 & -a_2\cos\theta_1\sin\theta_2 \\ 0 & -\cos\theta_2 & \sin\theta_2 & a_2\cos\theta_2 + d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{1} $$

Assume that the yawing angles of the two cameras are equal to α_1. It is known from Fig. 1 that the coordinates of point P in frame H are zero in the axes X_h and Y_h. The coordinate in the axis Z_h is

$$ z_h = D/(2\tan\alpha_1) \tag{2} $$

where D is the distance between the optical principal points of the two cameras, and α_1 is the yawing angle.

The position of point P in frame W can be calculated with (3) according to (1) and (2), i.e.,

$$ \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = {}^{w}T_h \begin{bmatrix} x_h \\ y_h \\ z_h \\ 1 \end{bmatrix} = \begin{bmatrix} -z_h\sin\theta_1\cos\theta_2 + a_2\sin\theta_1\sin\theta_2 \\ z_h\cos\theta_1\cos\theta_2 - a_2\cos\theta_1\sin\theta_2 \\ z_h\sin\theta_2 + a_2\cos\theta_2 + d_1 \\ 1 \end{bmatrix}. \tag{3} $$

C. General Bionic Visual Model

The general bionic visual model is designed for the robots without the neck. It consists of two cameras simultaneously

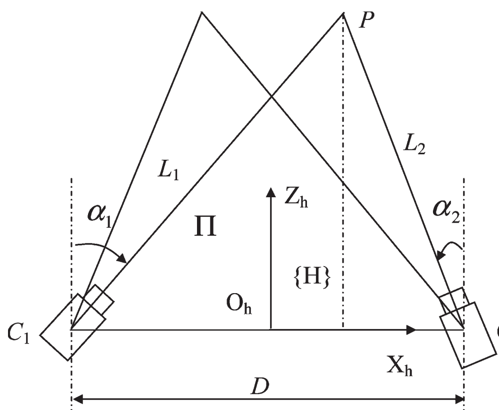


Fig. 4. Principle of visual positioning with the general model.

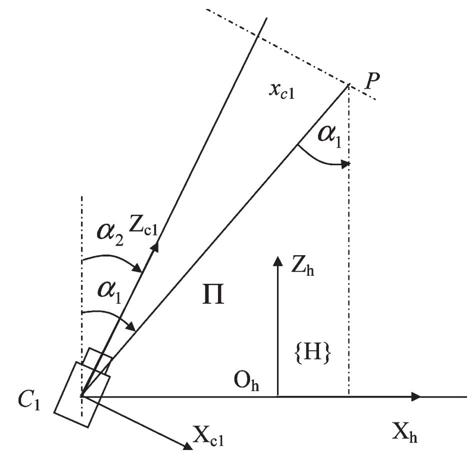


Fig. 5. Geometric relation for a camera.

yawing in opposite direction. In such a case, it is impossible to place the images of a point P at the center positions of the image planes of the two cameras at the same time. However, its horizontal imaging coordinates can be equal to those of the image plane centers of the two cameras separately. The cameras are simultaneously yawed in two steps, in which the coordinates of the image plane centers are taken as the desired values. In the first step, the horizontal imaging coordinate of point P in camera C_{a1} is adjusted to the desired value, and the image coordinates of point P in camera C_{a2} are recorded. In the second step, the horizontal imaging coordinate of point P in camera C_{a2} is adjusted to the desired value, and the image coordinates of point P in camera C_{a1} are recorded. The yawing angles in the two steps are recorded as α_1 and α_2. In the X_hZ_h plane, the geometric relation is shown in Fig. 4.

From the geometric relation in Fig. 4, z_h and x_h are computed as follows:

$$z_h = D/(\tan\alpha_1 + \tan\alpha_2) \tag{4}$$

$$x_h = z_h \tan\alpha_1 - D/2 \tag{5}$$

where α_1 is the yawing angle in the first step, and α_2 is the yawing angle in the second step.

For camera C_{a1}, the relation between the coordinates in image and Cartesian space can be expressed as follows according to the pinhole model with four intrinsic parameters:

$$\begin{cases} u_{11} - u_{10} = k_{x1}\frac{x_{c1}}{z_{c1}} \\ v_{11} - v_{10} = k_{y1}\frac{y_{c1}}{z_{c1}} \end{cases} \tag{6}$$

where $[u_{11}, v_{11}]$ are the image coordinates of point P in camera C_{a1} in the second step. $[u_{10}, v_{10}]$ are the image coordinates of the optical principal point, and u_{10} is used as the desired image coordinate in the first step. $[x_{c1}, y_{c1}, z_{c1}]$ are the Cartesian coordinates of point P in the frame of camera C_{a1} in the second step. k_{x1} and k_{y1} are the scale factors from imaging plane coordinates to the image coordinates.

y_{c1} can be deduced from (6) with the elimination of z_{c1}, i.e.,

$$y_{c1} = \frac{v_{11} - v_{10}}{u_{11} - u_{10}}\frac{k_{x1}}{k_{y1}}x_{c1} \approx \frac{v_{1d}}{u_{1d}}x_{c1} \tag{7}$$

where $u_{1d} = u_{11} - u_{10}$ and $v_{1d} = v_{11} - v_{10}$.

From the geometric relation as shown in Fig. 5, x_{c1} can be expressed with z_h, i.e.,

$$x_{c1} = \frac{\sin(\alpha_1 - \alpha_2)}{\cos\alpha_1}z_h \tag{8}$$

where α_1 and α_2 are same as described in (4).

Applying (8) to (7), y_{c1} can be obtained, i.e.,

$$y_{c1} \approx \frac{v_{1d}}{u_{1d}}\frac{\sin(\alpha_1 - \alpha_2)}{\cos\alpha_1}z_h. \tag{9}$$

Similarly, y_{c2} can be obtained as follows for camera C_{a2}:

$$y_{c2} \approx \frac{v_{2d}}{u_{2d}}\frac{\sin(\alpha_1 - \alpha_2)}{\cos\alpha_2}z_h \tag{10}$$

where $u_{2d} = u_{21} - u_{20}$ and $v_{2d} = v_{21} - v_{20}$. $[u_{21}, v_{21}]$ are the image coordinates of point P in camera C_{a2} in the first step. $[u_{20}, v_{20}]$ are the image coordinates of the optical principal point of camera C_{a2}, and u_{20} is used as the desired image coordinate of point P in the second step. y_{c2} is the Cartesian coordinate of point P on the Y_{c2}-axis in the frame of camera C_{a2} in the first step.

The average of y_{c1} and y_{c2} is taken as the coordinate y_h, i.e.,

$$y_h = (y_{c1} + y_{c2})/2. \tag{11}$$

The position of a point P in world frame W is easy to be obtained for the robot with a neck of two DOFs via coordinate transformation after its position in frame H is obtained [see also (3)]. This is very helpful for a robot to track an object in a large range.

III. Relative Positioning for Multiple Objects

Suppose that there are multiple objects in the common view field of two cameras. One object is selected as reference, and it is measured using the method in Section II-C. The symbols L_{11} and L_{12} denote optical lines in two steps for camera C_{a1}, and the symbols L_{21} and L_{22} for camera C_{a2}. The view fields can be divided into 12 areas from S_1 to S_{12}, as shown in Fig. 6, with lines L_{11}, L_{12}, L_{21}, and L_{22}, and the Z_h-axis. It is found that the areas S_1 and S_2 are distinguished with the Z_h-axis, so are

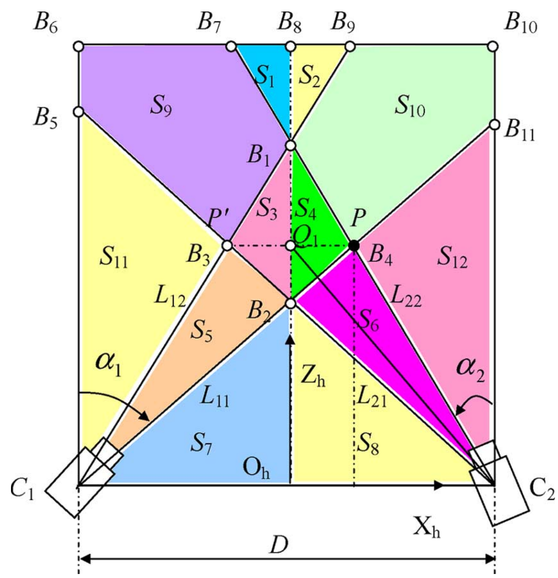


Fig. 6. Areas division in relative positioning.

the areas S_3 and S_4, and S_7 and S_8. The other areas are divided by optical lines L_{11}, L_{12}, L_{21}, and L_{22}.

Four frames of images are captured at the two measuring positions with yawing angles α_1 and α_2 for the two cameras. The image coordinates are indicated with $[u_{ijk}, v_{ijk}]$ for object k in the image j of camera i. The area in which object k locates can be determined with the image coordinates of object k and the optical principal points, i.e., $[u_{ijk}, v_{ijk}]$ and $[u_{i0}, v_{i0}]$, $i = 1, 2$, $j = 1, 2$. The division can be concluded as given in (12) from Fig. 6, i.e.,

$$S \in \begin{cases} S_1, & \text{if } u_{12k} < u_{10},\ u_{22k} > u_{20},\ |u_{12kd}| > |u_{22kd}| \\ S_2, & \text{if } u_{12k} < u_{10},\ u_{22k} > u_{20},\ |u_{12kd}| < |u_{22kd}| \\ S_3, & \text{if } u_{11k} < u_{10},\ u_{12k} > u_{10},\ u_{21k} > u_{20}, \\ & \quad u_{22k} < u_{20}, |u_{11kd}| > |u_{21kd}| \\ S_4, & \text{if } u_{11k} < u_{10},\ u_{12k} > u_{10},\ u_{21k} > u_{20}, \\ & \quad u_{22k} < u_{20}, |u_{11kd}| < |u_{21kd}| \\ S_5, & \text{if } u_{11k} < u_{10},\ u_{12k} > u_{10},\ u_{21k} < u_{20} \\ S_6, & \text{if } u_{11k} > u_{10},\ u_{21k} > u_{20},\ u_{22k} < u_{20} \\ S_7, & \text{if } u_{11k} > u_{10},\ u_{21k} < u_{20},\ |u_{11kd}| < |u_{21kd}| \\ S_8, & \text{if } u_{11k} > u_{10},\ u_{21k} < u_{20},\ |u_{11kd}| > |u_{21kd}| \\ S_9, & \text{if } u_{12k} < u_{10},\ u_{21k} > u_{20},\ u_{22k} < u_{20} \\ S_{10}, & \text{if } u_{11k} < u_{10},\ u_{12k} > u_{10},\ u_{22k} > u_{20} \\ S_{11}, & \text{if } u_{12k} < u_{10},\ u_{21k} < u_{20} \\ S_{12}, & \text{if } u_{11k} > u_{10},\ u_{22k} > u_{20} \end{cases} \tag{12}$$

where S is the area in which the object k locates. $u_{ijkd} = u_{ijk} - u_{i0}$.

After the area in which the object k locates is determined, the approximate position in the area can be estimated according to the image coordinates u_{ijk}. In addition, the areas S_3 and S_4 can be divided into subareas using auxiliary point Q_1, which is the intersection of line B_3B_4 and the Z_h-axis. The angle β is defined as $\angle B_2C_2Q_1$, which is given as follows:

$$\beta = \text{atan}(2z_h/D) + \alpha_1 - \pi/2. \tag{13}$$

The horizontal coordinate of point Q_1 in the first image of camera C_{a2} can be estimated as follows since it is in proportion to the imaging angle:

$$u_{21q} = u_{211}\beta/(\alpha_1 - \alpha_2) \tag{14}$$

where u_{21q} and u_{211} are the horizontal coordinates of point Q_1 and the reference object in the first image of camera C_{a2}.

Similarly, u_{12q}, the horizontal coordinate of point Q_1 in the second image of camera C_{a1}, can be estimated. Then, the areas such as S_3, S_4, S_5, S_6, S_9, and S_{10} can be further divided using u_{21q} and u_{12q}.

IV. Error Analysis

The error analysis is focused on the errors caused by the yawing mechanism for the two cameras.

For the model in Section II-B, the relative error can be calculated via the derivative of (2), i.e.,

$$dz_h/z_h = dD/D - 2d\alpha_1/\sin(2\alpha_1) \tag{15}$$

where dD is the error in D, and $d\alpha_1$ is the error in α_1.

Generally, $\alpha_1 \neq 0$. In the case of very little α_1, $\sin(2\alpha_1)$ will converge to $2\alpha_1$. Thus, (15) can be rewritten as

$$dz_h/z_h \approx dD/D - d\alpha_1/\alpha_1 \leq |dD/D| + |d\alpha_1/\alpha_1|. \tag{16}$$

From (16), it is easy to find that the relative error in z_h is proportional to relative errors dD/D and $d\alpha_1/\alpha_1$. For example, when the relative errors in D and α_1 are 1%, the relative error in z_h is not more than 2%.

For the model in Section II-C, the relative error can be calculated via the derivative of (4), i.e.,

$$\frac{dz_h}{z_h} = \frac{dD}{D} - \frac{(\cos\alpha_2/\cos\alpha_1)d\alpha_1 + (\cos\alpha_1/\cos\alpha_2)d\alpha_2}{\sin(\alpha_1 + \alpha_2)}. \tag{17}$$

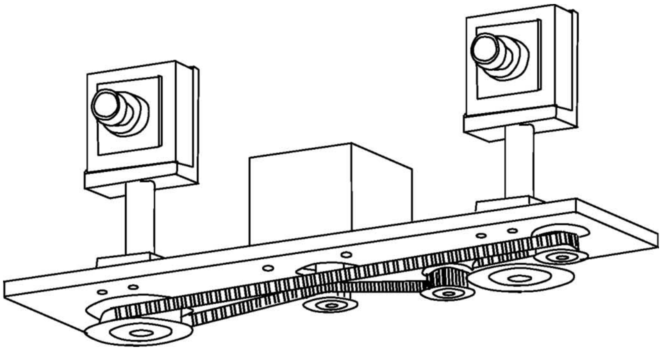
In general, $\alpha_1 > 0$ and $\alpha_2 > 0$; therefore, $\alpha_1 + \alpha_2 \neq 0$. If α_1 and α_2 are small enough, then (17) can be rewritten as follows:

$$\begin{aligned} dz_h/z_h &\approx dD/D - d(\alpha_1 + \alpha_2)/(\alpha_1 + \alpha_2) \\ &\leq |dD/D| + |d(\alpha_1 + \alpha_2)/(\alpha_1 + \alpha_2)|. \end{aligned} \tag{18}$$

If $d\alpha_1$ and $d\alpha_2$ are taken as the same, then (17) degenerates to (16).

The term $|d(\alpha_1 + \alpha_2)/(\alpha_1 + \alpha_1)|$ would be large if the errors $d\alpha_1$ and $d\alpha_2$ are large since the optical axes of the two cameras are not parallel in the initial state. In the initial state, the nonparallel axes can be taken as the results that the optical axes are yawed with initial angles. Hence, it is necessary to calibrate the initial angles of the optical axes relative to the Y_hZ_h plane. In fact, the influence of the principal point on the errors of z_h can be taken in the same way as for that of the initial angles and can be reduced via initial angle calibration.

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.



(a)

(b)

Fig. 7. Experiment system. (a) Principle scheme. (b) Actual system.

Fig. 8. Scene of calibration for initial optical directions.

From (5), the relative error dx_h/z_h is deduced as follows:

$$\frac{dx_h}{z_h} = \frac{dz_h}{z_h}\tan\alpha_1 - \frac{dD}{2z_h} + \frac{d\alpha_1}{(\cos\alpha_1)^2}. \tag{19}$$

In (19), the terms containing dD/z_h and $d\alpha_1$ are so small that they can be neglected. It is certain that dx_h/z_h is smaller than dz_h/z_h since $\tan\alpha_1 < 1$.

From (9) to (11), the relative error dy_h/z_h is deduced as follows:

$$\begin{aligned}\frac{dy_h}{z_h} = {} & \frac{1}{2}\frac{v_{1d}}{u_{1d}}\left[\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_1}\frac{dz_h}{z_h}\right.\\ & \left.\qquad + \frac{\cos\alpha_2 d\alpha_1 - \cos(\alpha_1-\alpha_2)\cos\alpha_1 d\alpha_2}{(\cos\alpha_1)^2}\right]\\ & + \frac{1}{2}\frac{v_{2d}}{u_{2d}}\left[\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_2}\frac{dz_h}{z_h}\right.\\ & \left.\qquad + \frac{\cos(\alpha_1-\alpha_2)\cos\alpha_2 d\alpha_1 - \cos\alpha_1 d\alpha_2}{(\cos\alpha_2)^2}\right]\\ & + \frac{1}{2}\frac{dv_{1d}}{u_{1d}}\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_1} + \frac{1}{2}\frac{dv_{2d}}{u_{2d}}\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_2}\\ & - \frac{1}{2}\frac{v_{1d}du_{1d}}{u_{1d}^2}\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_1} - \frac{1}{2}\frac{v_{2d}du_{2d}}{u_{2d}^2}\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_2}\end{aligned} \tag{20}$$

where du_{1d}, dv_{1d}, du_{2d}, and dv_{2d} are the errors in u_{1d}, v_{1d}, u_{2d}, and v_{2d}, respectively.

The terms such as $[\cos\alpha_2 d\alpha_1 - \cos(\alpha_1-\alpha_2)\cos\alpha_1 d\alpha_2]/(\cos\alpha_1)^2$ and $[\cos(\alpha_1-\alpha_2)\cos\alpha_2 d\alpha_1 - \cos\alpha_1 d\alpha_2]/(\cos\alpha_2)^2$ in (20) are negligible when the angles α_1 and α_2 are small enough. Terms with du_{1d} and du_{2d} are negligible after the initial angles of the optical axes are calibrated. Then, (20) can be rewritten as follows:

$$\begin{aligned}\frac{dy_h}{z_h} \approx {} & \frac{1}{2}\left[\frac{v_{1d}}{u_{1d}}\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_1} + \frac{v_{2d}}{u_{2d}}\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_2}\right]\frac{dz_h}{z_h}\\ & + \frac{1}{2}\left[\frac{dv_{1d}}{u_{1d}}\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_1} + \frac{dv_{2d}}{u_{2d}}\frac{\sin(\alpha_1-\alpha_2)}{\cos\alpha_2}\right].\end{aligned} \tag{21}$$

It is found from (21) that dy_h/z_h is smaller than dz_h/z_h since $\sin(\alpha_1-\alpha_2)/\cos\alpha_1 \ll 1$ and $\sin(\alpha_1-\alpha_2)/\cos\alpha_2 \ll 1$ when v_{1d} and v_{2d} are accurate, u_{1d} and u_{2d} are not very small, and α_1 and α_2 are small enough. In the case of very small u_{1d} and u_{2d}, the error dy_h/z_h will be large. An alternative method to solve this problem is given as follows. When y_{c1} is calculated with (9), u_{1d} and v_{1d} are generated in the condition $\alpha_2 = 0$. While y_{c2} is calculated with (10), u_{2d} and v_{2d} are generated in the condition $\alpha_1 = 0$. In the case that there are large errors in v_{1d} and v_{2d}, the error dy_h/z_h is apparent since it is proportional to dv_{1d} and dv_{2d}. In addition, k_x and k_y are very close for most cameras. Generally, the value of k_y/k_x is close to 1 with an error of less than 2%. For example, when $\alpha_1 = \pi/6$, $\alpha_2 = \pi/12$, $u_{1d} = 40$, $v_{1d} = 50$, $u_{2d} = 45$, $v_{2d} = 60$, $dz_h/z_h = 2\%$, $dv_{1d} = 50$, and $dv_{2d} = 50$, the relative error dy_h/z_h is not more than 1.1%. It means that the relative error dy_h/z_h is not very sensitive to the cameras' intrinsic parameters.

V. Calibrating the Initial Directions of the Optical Axes

From (16) and (18), it should be noted that the term dD/D is a small constant since $D \gg dD$. Thus, the relative errors in α_1 and α_2 may be the main source for the relative error in z_h.

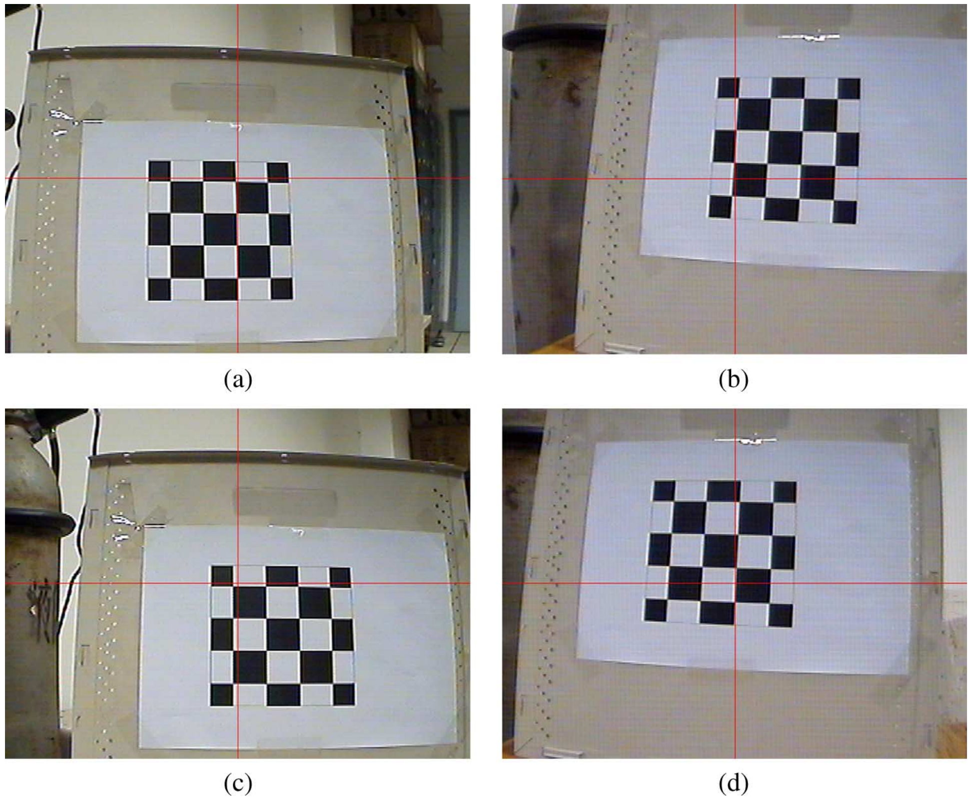


Fig. 9. Some images of the object to be measured in experiments. (a) Image of chessboard in C_{a1} and (b) image in C_{a2} at the first step. (c) Image in C_{a1} and (d) image in C_{a2} at the second step.

The initial yawing angles of the cameras are assumed to be zero, and the optical axes are assumed to be parallel. In fact, the actual initial yawing angles will not be zero. As mentioned in Section II-B, the optical axes of two cameras are just almost parallel in the initial state. Obviously, there exist system errors denoted as α_{e1} and α_{e2} for α_1 and α_2, respectively, in the initial state. The calibration of the initial directions of optical axes is to find the values of α_{e1} and α_{e2}.

Taking α_{e1} and α_{e2} into account, (4) is rewritten as follows:

$$\tan(\alpha_1 + \alpha_{e1}) + \tan(\alpha_2 + \alpha_{e2}) = D/z_h. \tag{22}$$

With the expansion and simplification of (22), the following equation is derived:

$$a_1 xy + a_2 x + a_3 y + a_4 = 0 \tag{23}$$

where

$$\begin{cases} x = \tan\alpha_{e1} \\ y = \tan\alpha_{e2} \\ a_1 = \tan\alpha_1 + \tan\alpha_2 + \tan\alpha_1 \tan\alpha_2 D/z_h \\ a_2 = \tan\alpha_1 \tan\alpha_2 - \tan\alpha_1 D/z_h - 1 \\ a_3 = \tan\alpha_1 \tan\alpha_2 - \tan\alpha_2 D/z_h - 1 \\ a_4 = D/z_h - \tan\alpha_1 - \tan\alpha_2. \end{cases} \tag{24}$$

Formula (23) is a nonlinear equation for parameters x and y. In the calibration, a block is placed in front of the two cameras; the distance from the block to the midpoint of the two cameras can be measured. The cameras are yawed to have α_1 and α_2 as described in Section II-C. Changing the block's position a number of times, a series of nonlinear equations as (23) are formed.

Let

$$f_i(x, y) = a_{1i} xy + a_{2i} x + a_{3i} y + a_{4i} \tag{25}$$

where a_{1i} to a_{4i} are the coefficients a_1 to a_4 computed from (24) at the ith sampling of calibrating data.

Then, an objective function $F(x, y)$ can be defined as follows:

$$F(x, y) = \sum_{i=1}^{n} f_i^2(x, y) \tag{26}$$

where n is the sampling times, i.e., the groups of data formed for calibration.

Now, the solution of the nonlinear (23) is converted to an optimization problem to find the optimal parameters x and y to make $F(x, y)$ be minimum. As it is known, the quasi-Newton method is efficient to solve this problem.

After the above calibration, the parameters u_{10} and u_{20} in (9) and (10) can be evaluated to the image horizontal coordinates of the image center.

VI. Experiments and Results

An experiment system was designed as shown in Fig. 7, in which Fig. 7(a) was its principle scheme, and Fig. 7(b) was the actual system. It consisted of two miniature cameras that could be simultaneously yawed in opposite directions. A step motor

TABLE I
MEASURED IMAGE OFFSET COORDINATES AND YAWING ANGLES

Points	u_{1d}, v_{1d}	u_{2d}, v_{2d}	α_1(rad)	α_2(rad)
1	-89, 3	91, -112	0.0633	0.1706
2	1, 4	-2, -112	0.1178	0.1171
3	89, 4	-91, -113	0.1721	0.0633
4	179, 5	-181, -114	0.2253	0.0103
5	-91, 47	95, -68	0.0631	0.1748
6	-2, 48	3, -68	0.1185	0.1212
7	88, 47	-92, -68	0.1740	0.0663
8	179, 49	-181, -68	0.2263	0.0113
9	-95, 91	98, -23	0.0626	0.1777
10	-4, 91	5, -24	0.1195	0.1235
11	87, 91	-90, -22	0.1752	0.0692
12	178, 91	-181, -21	0.2278	0.0140
13	-98, 137	100, 22	0.0633	0.1811
14	-7, 137	8, 24	0.1195	0.1274
15	85, 136	-88, 25	0.1757	0.0714
16	176, 136	-181, 25	0.2297	0.0167

TABLE II
MEASURED IMAGE OFFSET COORDINATES FOR SOME POINTS TO CALCULATE y

Points	u_{1d}, v_{1d}	u_{2d}, v_{2d}
2	99, 4	-106, -115
6	99, 46	-109, -70
10	98, 91	-112, -25
14	99, 136	-114, 22

was employed to drive the rotation of cameras through the belt and gears. The system was adjusted so that the optical axes of the two cameras were almost parallel initially. The distance between the two cameras was 150 mm. The rotational resolution of the two cameras was $2\pi/25\,600 = 2.45 \times 10^{-4}$ rad.

A series of measurement experiments were conducted with the visual system, as shown in Fig. 7(b). First, the initial directions of the optical axes of two cameras were calibrated with the method described in Section V. A scene of optical initial direction calibration was given in Fig. 8. The results were $\alpha_{e1} = 0.0578$ rad and $\alpha_{e2} = -0.0254$ rad. Then, the measurement method, as described in Section II-C, was employed in the visual measuring experiments.

A. Chessboard Measurement

An experiment to measure the blocks in a chessboard was designed to test the effectiveness of the proposed method and system. In the visual measuring experiment, the cameras were yawed to make the horizontal imaging coordinates of the feature point be equal to those of the image plane centers of the two cameras separately for each point to be measured in Cartesian space. As described in Section II-C, the cameras were yawed in two steps, and two yawing angles α_1 and α_2 were generated. In Fig. 9, the images captured by the two cameras for the measure of a point were given. Fig. 9(a) was an image of chessboard in C_{a1}, Fig. 9(b) an image in C_{a2} at the first step, Fig. 9(c) an image in C_{a1}, and Fig. 9(d) an image in C_{a2} at the second step. The image size was 640×480 in pixel, and its center was [320, 240]. In the experiment, u_{10} and u_{20} were evaluated to be 320; v_{10} and v_{20} were evaluated to be 240. It can be seen that the images have large distortions, and the optical axes of the two cameras might not be parallel.

TABLE III
MEASURED RESULTS IN 3-D POSITIONS FOR THE CROSS POINTS ON A CHESSBOARD

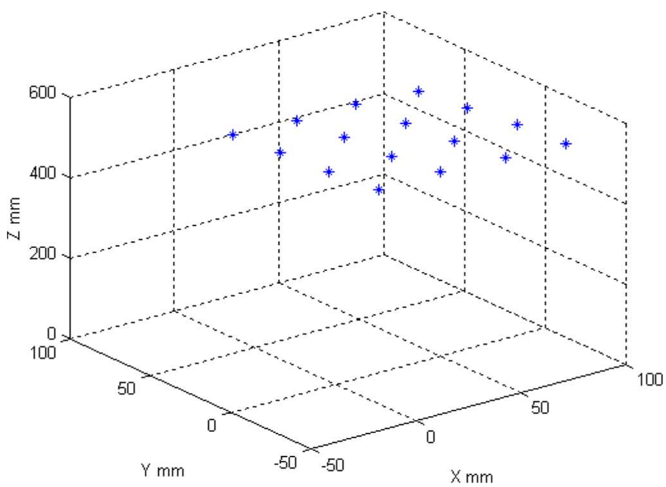
Points	X (mm)	Y (mm)	Z (mm)
1	-6.8551	-36.2563	559.8564
2	23.8075	-34.0894	556.8579
3	54.0711	-35.8102	551.6178
4	83.2076	-34.8537	543.8075
5	-7.9986	-6.2567	551.5902
6	22.4331	-6.0727	546.7718
7	52.8529	-5.5617	541.6156
8	82.6170	-5.2144	539.8013
9	-8.8849	22.8316	546.5353
10	21.8423	22.9313	540.3840
11	51.6135	23.4384	533.4364
12	81.0656	23.5013	531.5773
13	-9.4947	51.5836	538.1701
14	20.4779	50.8963	532.7708
15	50.6692	53.0560	528.3036
16	79.5523	52.1520	522.6241

The image offset coordinates from the image center and the yawing angles for cross points in the chessboard were listed in Table I. It can be seen that the offset coordinates u_{1d} and u_{2d} of points 2, 6, 10, and 14 were very small. As analyzed at the end of Section IV, the calculation of y_{c1} and y_{c2} would introduce large errors. To deal with this problem, several data were captured for the four points above, that is, u_{1d} and v_{1d} were generated in the condition $\alpha_2 = 0$, and u_{2d} and v_{2d} were generated in the condition $\alpha_1 = 0$.

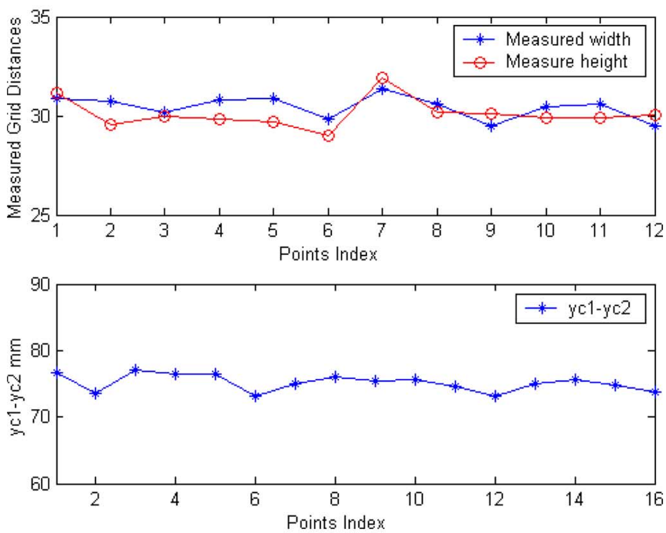
The coordinates z_h and x_h in frame H were computed using (4) and (5) according to α_1 and α_2 modified with α_{e1} and α_{e2}. With the image coordinates and yawing angles listed in Table I, y_{c1} and y_{c2} were calculated via (9) and (10), except for points 2, 6, 10, and 14. It should be noted that the term $\alpha_1 - \alpha_2$ in (9) and (10) denoted the relative rotation angle. Thus, α_1 and α_2 in the numerators of (9) and (10) did not need to be modified with α_{e1} and α_{e2}. α_1 in the denominator of (9) and α_2 in the denominator of (10) were the yawing angles relative to the axis Z_h, and they need to be modified with α_{e1} and α_{e2}. For points 2, 6, 10, and 14, y_{c1} was calculated via (9) with the image offset coordinates in Table II, α_1 in Table I, and $\alpha_2 = 0$. y_{c2} was calculated for these points via (10) with the image offset coordinates in Table II, α_2 in Table I, and $\alpha_1 = 0$. The average value of y_{c1} and y_{c2} was taken as the coordinate y_h. The experimental results to measure a chessboard were listed in Table III. The data were the 3-D positions of the cross points on the chessboard in the vision system frame H. They were also shown in Fig. 10(a) for convenience of evaluation. The actual width and height for each block in the chessboard were both 30 mm. The measured width and height computed from the distances between any two adjacent cross points in the pattern were listed in Table IV and also shown in Fig. 10(b). Its mean is 30.3 mm, and the standard deviation was 0.677 mm. In addition, Fig. 10(b) also displayed the difference of y_{c1} and y_{c2} computed from (9) and (10). It can be found that the differences were stable. Therefore, the differences can be considered as the offsets resulting from the nonparallel axes of the two cameras, in respect of an object in some depth Z_h.

From Fig. 10 and Tables III and IV, it can be found that the measuring accuracy with the proposed visual system and

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.



(a)



(b)

Fig. 10. Experiment results. (a) Measured results of cross points of a chessboard. (b) Measured width and height of the blocks in the chessboard, and the difference between y_{c1} and y_{c2}.

TABLE IV
MEASURED WIDTH AND HEIGHT OF THE BLOCKS ON A CHESSBOARD

No.	1	2	3	4	5	6	7	8	9	10	11	12
Width (mm)	30.9	30.6	30.2	30.8	30.9	29.8	31.3	30.6	29.5	30.5	30.6	29.5
Height (mm)	31.1	29.5	30.0	29.8	29.7	29.0	31.9	30.2	30.1	29.9	29.9	30.1

method was satisfactory even if the camera lens had large distortion.

B. Comparison With Stereovision

To compare the proposed method with the traditional stereovision method, the two cameras were well calibrated with Zhang's calibration method [6]. The intrinsic parameters of the cameras were as follows: $k_{x1} = 834.82771$, $k_{y1} = 815.41740$, $u_{10} = 303.8$, $v_{10} = 306.3$, $k_{x2} = 850.45548$, $k_{y2} = 833.29453$, $u_{20} = 345.1$, and $v_{20} = 197.3$. The distortion factors of the lens in the radial direction were $k_{c1} = -0.38741$ and $k_{c2} = -0.30938$ for cameras C_{a1} and C_{a2} separately. The extrinsic parameter matrix ${}^{c1}T_{c2}$, i.e., the

TABLE V
MEASURED POSITIONS WITH THE STEREOVISION METHOD AND THE PROPOSED METHOD USING THE PRINCIPAL POINT

Index	Proposed method (v_{10}=306, v_{20}=197)			Stereovision method		
	X(mm)	Y(mm)	Z(mm)	X(mm)	Y(mm)	Z(mm)
1	-34.1292	16.1916	539.3316	-36.1230	15.7195	540.0396
2	-32.2241	15.7743	693.0157	-34.9313	15.5160	696.7671
3	-49.7623	16.1721	827.8718	-53.2925	15.9535	830.8421
4	-48.2818	5.5954	1065.2693	-51.5189	5.3242	1032.9899
5	-67.6102	6.2453	1196.2755	-69.4528	6.0720	1178.5692
6	-83.6350	6.8764	1361.9719	-87.9713	6.1010	1341.4960

TABLE VI
MEASURED POSITIONS WITH THE PROPOSED METHOD IN THE CASE OF USING IMAGE CENTER AS THE PRINCIPAL POINT

Index	Proposed method (v_{10}=240, v_{20}=240)		
	X(mm)	Y(mm)	Z(mm)
1	-34.1292	26.1603	539.3316
2	-32.2241	26.2009	693.0157
3	-49.7623	26.5460	827.8718
4	-48.2818	18.3452	1065.2693
5	-67.6102	23.3014	1196.2755
6	-83.6350	26.4331	1361.9719

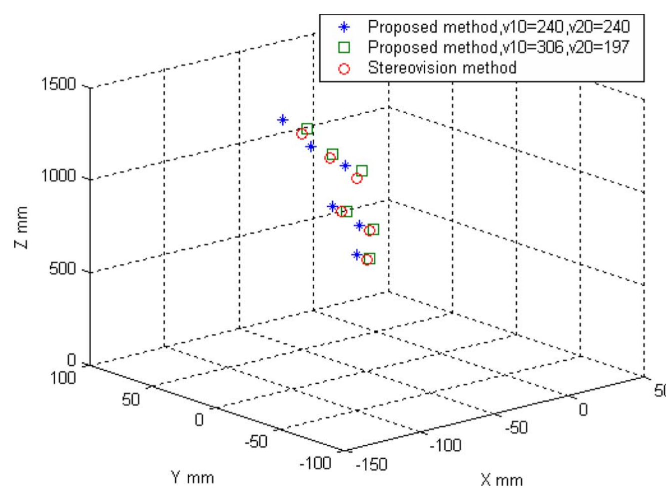
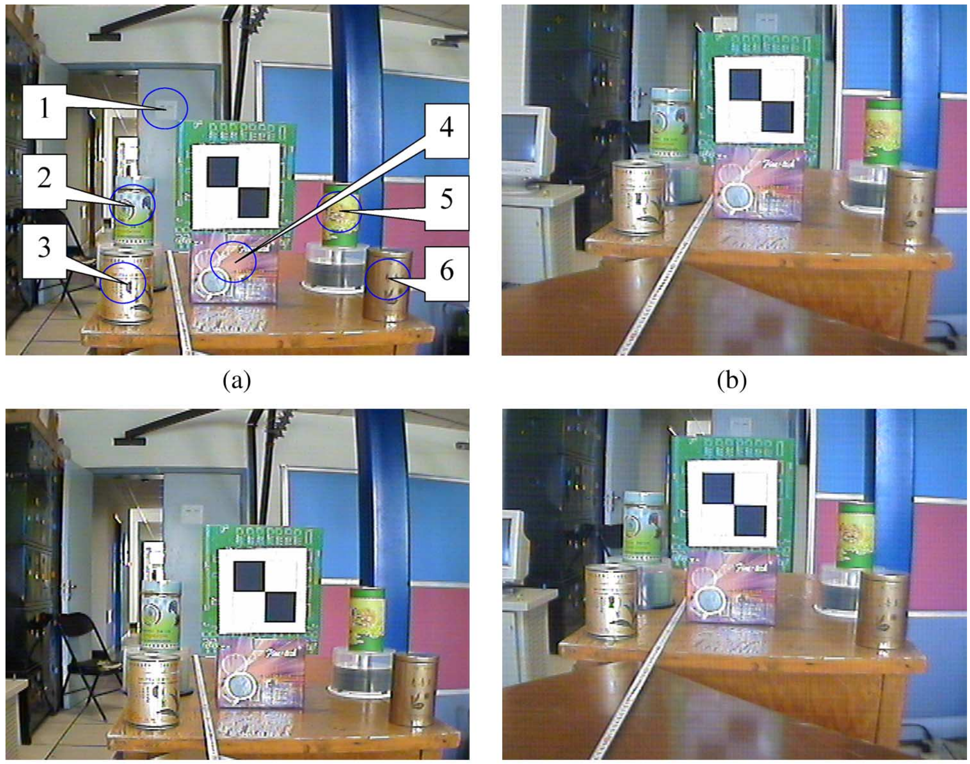


Fig. 11. Experiment results with the proposed method and the stereovision method.

pose of camera C_{a2} relative to camera C_{a1}, was well calibrated as given in (27) when the two cameras were at the initial positions, i.e.,

$$ {}^{c1}T_{c2} = \begin{bmatrix} 0.9995 & -0.0236 & -0.0222 & -150.9556 \\ 0.0234 & 0.9997 & -0.0091 & -5.1851 \\ 0.0224 & 0.0086 & 0.9997 & 2.2226 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{27} $$

The experiment scene was similar to that of the initial optical direction calibration, as given in Fig. 8. The intersection between the two black blocks on a target, as shown in Fig. 8, was selected as the point P to be measured. When the target was placed at a position in front of the visual system, the two cameras were yawed to initial directions and captured the target's images. The Cartesian space position of point P in the frame



(a) (b)

(c) (d)

Fig. 12. Images of the objects in experiments. (a) Image in C_{a1} and (b) image in C_{a2} at the first step. (c) Image in C_{a1} and (d) image in C_{a2} at the second step.

of camera C_{a1} was calculated with the traditional stereovision method. The coordinates of point P in frame H were obtained via transformations including the rotation with α_{e1} around axis y_{c1} and the translation with $D/2$ along axis x_{c1}. Then, the two cameras were yawed with a tracking algorithm in two steps to generate α_1 and α_2, and the coordinates of point P in frame H were computed with the proposed method as described in Section II-C. The procedure above was repeated while the target was placed at different positions in front of the visual system, and six groups of visual measuring results were formed as given in Tables V and VI. They are also displayed in Fig. 11 for assessing convenience.

The results from the stereovision method were computed using the intrinsic and extrinsic parameters of the two cameras, as given above in this section. The lens distortion in the radial direction was also taken into account. The results from the proposed method did not use the parameters such as k_{x1}, k_{y1}, k_{x2}, and k_{y2}, and the distortion factors k_{c1} and k_{c2}. The y-coordinates of the measured positions with the proposed method in Table V were computed in the condition that $u_{10} = 320$, $v_{10} = 306.3$, $u_{20} = 320$, and $v_{20} = 197.3$. The y-coordinates in Table VI were computed with the proposed method in the condition that $u_{10} = 320$, $v_{10} = 240$, $u_{20} = 320$, and $v_{20} = 240$. In other words, the results in Table VI were calculated in the case that the intrinsic parameters of the cameras were supposed to be not available.

From Fig. 11 and Tables V and VI, it can be found that the measuring accuracy with the proposed visual system and method was very close to that with the stereovision method, even if the cameras' intrinsic parameters were not employed, and the large distortion in the camera lens was not taken into account in the proposed method.

C. Relative Positioning

To verify the effectiveness of the relative positioning method for multiple objects, an experiment was conducted. A board target with two black blocks, as shown in Fig. 8, was selected as the main object, which was surrounded by other objects. The intersection of the two blocks was selected as the feature point. As described in Section II-C, the cameras were yawed with a tracking algorithm in two steps. In the first step, the cameras were yawed to make the feature point be at the horizontal center in the image of camera C_{a1}. In the second step, the cameras were yawed to make the feature point be at the horizontal center in the image of camera C_{a2}. α_1 and α_2 were generated as $\alpha_1 = 0.08$ and $\alpha_2 = 0.0349$. Each camera captured an image at the end of each step. Four frames of images were captured at the two measuring positions for the two cameras, as shown in Fig. 12.

The six objects to be measured were represented by their image centers. The image coordinates u_{11k}, u_{12k}, u_{21k}, and u_{22k}, $k = 1, 2, \ldots, 6$, for the six objects extracted from the four images captured by the two cameras in the two steps, were listed in Table VII. Applying (12) to the image coordinates of the six objects, we had the areas that the objects belonged to. In other words, the approximate positions of the objects relative to the main object were obtained, as listed in Table VII. It is easy to check the correctness of the relative positioning results via comparison to their actual positions.

TABLE VII
IMAGE COORDINATES OF THE OBJECTS AND THEIR AREAS LOCATED

Object k	u_{11k}	u_{12k}	u_{22k}	u_{22k}	Area belonged to Measured	Actual
1	220	256	363	327	S_1	S_1, left, behind
2	177	213	227	190	S_{11}	S_{11}, left
3	165	202	186	148	S_{11}	S_{11}, left
4	314	353	352	317	S_4	S_4, left, front
5	463	500	527	490	S_{12}	S_{12}, right
6	532	569	560	525	S_{12}	S_{12}, right

In addition, experiments with the proposed visual system and method, in Sections VI-A and B, also gave evidence that the measuring precision would be heavily influenced by the directions of the optical axes of the two cameras in the initial state. Therefore, the calibration of the initial directions of the optical axes of the two cameras is important to ensure the precision in practical visual measurements.

VII. CONCLUSION

A new active visual system is developed, which consists of two cameras and a two-DOF mechanical platform. Two cameras are mounted on the platform, which can pitch and yaw. The two cameras can be simultaneously adjusted in opposite directions. With pitching and yawing of the platform, and relative yawing of the cameras, the object's images can be adjusted to the center areas of the image planes of the two cameras. Then, the position of the object is determined with the geometrical information of the visual system. Furthermore, a more general visual model is proposed. It consists of two cameras that can yaw in opposite directions. In two steps, the object's images are adjusted to the center areas of the image planes of the two cameras separately. The position of an object can be calculated with the yawing angles and the image coordinates of the object in the two steps.

The visual system proposed in this paper is based on bionic vision and is insensitive to the intrinsic parameters of the camera. Experiment results showed that the measuring accuracy with the proposed visual system and method was very close to that with a stereovision method, even if the actual intrinsic parameters of the cameras were not available, and large distortion in the camera lens was not taken into account in the proposed method. Low efficiency in measuring multiple objects is its main limitation. However, the cases with the tracking or measuring of multiple objects are uncommon in a visual control system.

Future work will be focused on its applications such as navigation, object tracking, approaching, and grasping for humanoid robots.

REFERENCES

[1] G. D. Hager, S. Hutchinson, and P. I. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 651–670, Oct. 1996.
[2] J. G. Juang, "Parameter estimation in the three-point perspective projection problem in computer vision," in *Proc. IEEE Int. Symp. Ind. Electron.*, Jul. 1997, vol. 3, pp. 1065–1070.
[3] D. Xu, Y. F. Li, and M. Tan, "A visual positioning method based on relative orientation detection for mobile robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Beijing, China, Oct. 9–15, 2006, pp. 1243–1248.
[4] A. Sugimoto, W. Nagatomo, and T. Matsuyama, "Estimating ego motion by fixation control of mounted active cameras," in *Proc. Asian Conf. Comput. Vis.*, 2004, vol. 1, pp. 67–72.
[5] O. D. Faugeras and G. Toscani, "The calibration problem for stereo," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 1986, pp. 15–20.
[6] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
[7] D. Xu, Y. F. Li, and M. Tan, "Method for calibrating cameras with large distortion in lens," *Opt. Eng.*, vol. 45, no. 4, p. 043602, Apr. 2006.
[8] J. Qian and J. Su, "Online estimation of image Jacobian matrix by Kalman–Bucy filter for uncalibrated stereo vision feedback," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2002, vol. 1, pp. 562–567.
[9] E. Malis, F. Chaumette, and S. Boudet, "2 1/2D visual servoing," *IEEE Trans. Robot. Autom.*, vol. 15, no. 2, pp. 238–250, Apr. 1999.
[10] S. D. Ma, "A self-calibration technique for active vision system," *IEEE Trans. Robot. Autom.*, vol. 12, no. 1, pp. 114–120, Feb. 1996.
[11] E. Guillou, D. Meneveaux, E. Maisel, and K. Bouatouch, "Using vanishing points for camera calibration and coarse 3D reconstruction from a single image," *Vis. Comput.*, vol. 16, no. 7, pp. 396–410, 2000.
[12] A. Almansa and A. Desolneux, "Vanishing point detection without any *a priori* information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 4, pp. 502–507, Apr. 2003.
[13] D. Xu, Y. F. Li, Y. Shen, and M. Tan, "New pose detection method for self-calibrated cameras based on parallel lines and its application in visual control system," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 5, pp. 1104–1117, Oct. 2006.
[14] D. Kragic, A. T. Miller, and P. K. Allen, "Real-time tracking meets online grasp planning," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2001, vol. 3, pp. 2460–2465.
[15] J. A. Piepmeier, G. V. McMurray, and H. Lipkin, "Uncalibrated dynamic visual servoing," *IEEE Trans. Robot. Autom.*, vol. 20, no. 1, pp. 143–147, Feb. 2004.
[16] Y. Shen, G. Xiang, Y.-H. Liu, and K. Li, "Uncalibrated visual servoing of planar robots," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2002, vol. 1, pp. 580–585.
[17] C. E. Smith and N. P. Papanikolopoulos, "Grasping of static and moving objects using a vision-based control approach," *J. Intell. Robot. Syst.: Theory Appl.*, vol. 19, no. 3, pp. 237–270, 1997.
[18] D. Xu, M. Tan, and Y. Shen, "A new simple visual control method based on cross ratio invariance," in *Proc. IEEE Int. Conf. Mechatronics Autom.*, Niagara Falls, ON, Canada, Jul. 29–Aug. 1 2005, pp. 370–375.
[19] Y. Shen, D. Xu, and M. Tan, "Torch transferring between two humanoid robots with the guidance of visual information," in *Proc. SICE Annu. Conf., Int. Conf. Instrum., Control Inf. Technol.*, Okayama, Japan, Aug. 8–10, 2005, pp. 510–515.
[20] H. Bie, Q. Huang, W. Zhang, B. Song, and K. Li, "Visual tracking of a moving object of a robot head with 3 DOF," in *Proc. IEEE Int. Conf. Robot., Intell. Syst. Signal Process.*, 2003, vol. 1, pp. 686–691.

De Xu (M'05) received the B.Sc. and M.Sc. degrees from the Shandong University of Technology, Jinan, China, in 1985 and 1990, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2001, all in control science and engineering.

He was a Senior Researcher Associate from March 1, 2005 to September 1, 2005 and a Research Fellow from January 22, 2007 to March 21, 2007 with the City University of Hong Kong, Kowloon. Since 2001, he has been with the Institute of Automation, Chinese Academy of Sciences (IACAS), Beijing. He is currently a Professor with the Laboratory of Complex Systems and Intelligence Science, IACAS. His research interests include robotics and automation, particularly the control of robots such as visual control and intelligent control.

You Fu Li (S'91–M'92–A'95–SM'01) received the B.Sc. and M.Sc. degrees in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 1982 and 1986, respectively, and the Ph.D. degree in robotics from Oxford University, Oxford, U.K., in 1993.

From 1993 to 1995, he was a Postdoctoral Researcher with the Department of Computer Science, University of Wales, Aberystwyth, U.K. In 1995, he joined the City University of Hong Kong, Kowloon. He has published over 100 papers in refereed international journals and conferences. His research interests include robot vision, visual tracking, robot sensing and sensor-based control, mechatronics, and automation.

Dr. Li is an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING.

Min Tan received the B.Sc. degree from Tsing Hua University, Beijing, China, in 1986 and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences (IACAS), Beijing, in 1990, both in control science and engineering.

He is currently a Professor with the Laboratory of Complex Systems and Intelligence Science, IACAS. He has published over 100 papers in journals, books, and conferences. His research interests include robotic control and intelligent control.

Yang Shen received the B.Sc. degree in control science and engineering from the University of Science and Technology of China, Hefei, in 2002. He is currently working toward the Ph.D. degree at the Institute of Automation, Chinese Academy of Sciences, Beijing.

His research interests include robotics and automation.