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.

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 camera’s 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...
is on the plane formed by the two optical axes. The following
rections to stare at an object. In the initial state of the two
iments, 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


\[ z_h = D/(2 \tan \alpha_1) \]  

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

\[ \begin{bmatrix}
  x_w \\
  y_w \\
  z_w
\end{bmatrix} = w \begin{bmatrix}
  x_h \\
  y_h \\
  z_h
\end{bmatrix} = \begin{bmatrix}
  -z_h \sin \theta_1 \cos \theta_2 + a_2 \sin \theta_1 \sin \theta_2 \\
  z_h \cos \theta_1 \cos \theta_2 - a_2 \cos \theta_1 \sin \theta_2 \\
  z_h \sin \theta_2 + a_2 \cos \theta_2 + d_1
\end{bmatrix} \]  

C. General Bionic Visual Model

The general bionic visual model is designed for the robots

The two cameras can simultaneously yaw in opposite di-
rections. According to (1) and (2), i.e.,

\[ \begin{bmatrix}
  x_w \\
  y_w \\
  z_w
\end{bmatrix} = w \begin{bmatrix}
  x_h \\
  y_h \\
  z_h
\end{bmatrix} = \begin{bmatrix}
  -z_h \sin \theta_1 \cos \theta_2 + a_2 \sin \theta_1 \sin \theta_2 \\
  z_h \cos \theta_1 \cos \theta_2 - a_2 \cos \theta_1 \sin \theta_2 \\
  z_h \sin \theta_2 + a_2 \cos \theta_2 + d_1
\end{bmatrix}. \]  

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 \( \Pi \) optical principal point, respectively, of another camera \( C_{a2} \). \( \Pi \) 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 \( W \).

For a point \( P \), it can be adjusted to be on the plane \( \Pi \) with the change in \( \theta_2 \). Then, it can be on the perpendicular bisector of line \( C_1 C_2 \) on the plane \( \Pi \) with the adjustment of \( \theta_1 \). With simultaneous yawing in opposite directions for the two cameras, the images of point \( P \) can be placed at the center 126 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, the two cameras, and \( \theta_2 \) are the joint angles of the two joints.

\[ wT_h = \begin{bmatrix}
  \cos \theta_1 & -\sin \theta_1 & 0 & a_2 \sin \theta_1 \sin \theta_2 \\
  \sin \theta_1 & \cos \theta_1 & 0 & a_2 \cos \theta_1 \sin \theta_2 \\
  0 & 0 & 1 & a_2 \cos \theta_2 + d_1
\end{bmatrix}. \]  

Assume that the yawing angles of the two cameras are equal to \( \alpha_1 \). 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 134 axis \( Z_h \) is

\[ z_h = D/(2 \tan \alpha_1) \]  

where \( D \) is the distance between the optical principal points of 136 the two cameras, and \( \alpha_1 \) is the yawing angle.

The position of point \( P \) in frame \( W \) can be calculated with 138 (3) according to (1) and (2), i.e.,

\[ \begin{bmatrix}
  x_w \\
  y_w \\
  z_w
\end{bmatrix} = \begin{bmatrix}
  x_h \\
  y_h \\
  z_h
\end{bmatrix} = \begin{bmatrix}
  -z_h \sin \theta_1 \cos \theta_2 + a_2 \sin \theta_1 \sin \theta_2 \\
  z_h \cos \theta_1 \cos \theta_2 - a_2 \cos \theta_1 \sin \theta_2 \\
  z_h \sin \theta_2 + a_2 \cos \theta_2 + d_1
\end{bmatrix}. \]  

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 di-
rections to stare at an object. In the initial state of the two
images, 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. 1. Structure of a humanoid robot.

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

II. BIONIC VISUAL MODEL

A. Visual System for a Humanoid Robot

B. Bionic Visual Model for a Humanoid Robot
yawing in opposite direction. In such a case, it is impossible
144 to place the images of a point \( P \) at the center positions of the
145 image planes of the two cameras at the same time. However,
146 its horizontal imaging coordinates can be equal to those of
147 the image plane centers of the two cameras separately. The
148 cameras are simultaneously yawed in two steps, in which the
149 coordinates of the image plane centers are taken as the desired
150 values. In the first step, the horizontal imaging coordinate of point \( P \) in camera \( C_{a1} \) is adjusted to the desired value, and the
151 image coordinates of point \( P \) in camera \( C_{a2} \) are recorded. In
152 the second step, the horizontal imaging coordinate of point \( P \) in camera \( C_{a2} \) is adjusted to the desired value, and the image
153 coordinates of point \( P \) in camera \( C_{a1} \) are recorded. The yawing
154 angles in the two steps are recorded as \( \alpha_1 \) and \( \alpha_2 \). In the \( X_hZ_h\)
155 plane, the geometric relation is shown in Fig. 4.

158 From the geometric relation in Fig. 4, \( z_h \) and \( x_h \) are com-
159 puted as follows:

\[
z_h = D / (\tan \alpha_1 + \tan \alpha_2)
\]
\[
x_h = z_h \tan \alpha_1 - D / 2
\]

160 where \( \alpha_1 \) is the yawing angle in the first step, and \( \alpha_2 \) is the
161 yawing angle in the second step.

162 For camera \( C_{a1} \), the relation between the coordinates in im-
163 age and Cartesian space can be expressed as follows according
164 to the pinhole model with four intrinsic parameters:

\[
\begin{align*}
u_{11} - u_{10} &= k_{x1} x_{c1} \\
v_{11} - v_{10} &= k_{y1} y_{c1}
\end{align*}
\]

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

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

\[
y_{c1} = \frac{v_{11} - v_{10}}{u_{11} - u_{10}} k_{x1} x_{c1} \approx \frac{v_{1d}}{u_{1d}} x_{c1}
\]

173 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
174 expressed with \( z_h \), i.e.,

\[x_{c1} = \frac{\sin(\alpha_1 - \alpha_2)}{\cos \alpha_1} z_h\]

175 where \( \alpha_1 \) and \( \alpha_2 \) are same as described in (4).

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

\[y_{c1} = v_{1d} \sin(\alpha_1 - \alpha_2) \frac{\cos \alpha_1}{z_h} z_h.
\]

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

\[y_{c2} = v_{2d} \sin(\alpha_1 - \alpha_2) \frac{\cos \alpha_2}{z_h} z_h.
\]

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
181 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
182 coordinate of point \( P \) on the \( Y_{c2} \)-axis in the frame of camera
183 \( C_{a2} \) in the first step.

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

\[y_h = (y_{c1} + y_{c2}) / 2.
\]

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

III. RELATIVE POSITIONING FOR MULTIPLE OBJECTS

192 Suppose that there are multiple objects in the common view
193 field of two cameras. One object is selected as reference, and it is
194 measured using the method in Section II-C. The symbols \( L_{11} \) and
195 \( L_{12} \) denote optical lines in two steps for camera \( C_{a1} \), and
196 the symbols \( L_{21} \) and \( L_{22} \) for camera \( C_{a2} \). The view fields can be
197 divided into 12 areas from \( S_1 \) to \( S_{12} \), as shown in Fig. 6, with
198 lines \( L_{11}, L_{12}, L_{21}, \) and \( L_{22} \), and the \( Z_h \)-axis. It is found that
199 the areas \( S_1 \) and \( S_2 \) are distinguished with the \( Z_h \)-axis, so are 200
the areas $S_1$ and $S_4$, and $S_7$ and $S_6$. 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 $\alpha_1$ and $\alpha_2$ 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., $[u_{i0}, v_{i0}]$. When the relative errors in $\alpha$ are small enough, then (17) can be rewritten as

$$\frac{dz_h}{z_h} \approx \frac{dD}{D} - \frac{d\alpha_1}{\alpha_1} \leq |\frac{dD}{D}| + |\frac{d\alpha_1}{\alpha_1}|.$$  (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, 237 when the relative errors in $D$ and $\alpha_1$ 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.,

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

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

$$\frac{dz_h}{z_h} \approx \frac{dD}{D} - \frac{d(\alpha_1 + \alpha_2)}{(\alpha_1 + \alpha_2)} \leq |\frac{dD}{D}| + |\frac{d(\alpha_1 + \alpha_2)}{(\alpha_1 + \alpha_2)}|.$$  (18)

If $d\alpha_1$ and $d\alpha_2$ are taken as the same, then (17) degenerates 245 to (16).

The term $|d(\alpha_1 + \alpha_2)/(\alpha_1 + \alpha_2)|$ would be large if the 247 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 250 are yawed with initial angles. Hence, it is necessary to calibrate 251 the initial angles of the optical axes relative to the $Y_hZ_h$ plane. 252 In fact, the influence of the principal point on the errors of $z_h$ 253 can be taken in the same way as for that of the initial angles and 254 can be reduced via initial angle calibration.
From (9) to (11), the relative error $dy_h/z_h$ is deduced as follows:

$$
\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} + \frac{\cos \alpha_2 \alpha_1 - \cos(\alpha_1 - \alpha_2) \cos \alpha_2 \alpha_1}{(\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} + \frac{\cos(\alpha_1 - \alpha_2) \cos \alpha_2 \alpha_1 - \cos \alpha_2 \alpha_1}{(\cos \alpha_2)^2} \right] + \frac{1}{2} \frac{v_{1d}}{u_{1d}} \sin(\alpha_1 - \alpha_2) + \frac{1}{2} \frac{v_{2d}}{u_{2d}} \sin(\alpha_1 - \alpha_2) - \frac{1}{2} \frac{v_{1d}u_{1d}}{u_{2d} \cos \alpha_1} \sin(\alpha_1 - \alpha_2) - \frac{v_{2d}u_{2d}}{u_{2d} \cos \alpha_2} \sin(\alpha_1 - \alpha_2)

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

$$
\frac{dy_h}{z_h} \approx \frac{1}{2} \left[ \frac{v_{1d}}{u_{1d}} \sin(\alpha_1 - \alpha_2) + \frac{v_{2d}}{u_{2d}} \sin(\alpha_1 - \alpha_2) \right] \frac{dz_h}{z_h} + \frac{1}{2} \left[ \frac{du_{1d}}{u_{1d}} \sin(\alpha_1 - \alpha_2) + \frac{dv_{2d}}{u_{2d}} \sin(\alpha_1 - \alpha_2) \right].

(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 $\alpha_1$ and $\alpha_2$ 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 (25) and $u_{1d}$ and $v_{1d}$ are generated in the condition $\alpha_2 = 0$. While $u_{2d}$ and $v_{2d}$ are 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 $du_{1d}$ and $dv_{2d}$. In addition, $k_x$ and $k_y$ are very close for most cameras. Generally, the value of $k_x/k_y$ 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 $\alpha_1$ and $\alpha_2$ may be the main source for the relative error in $z_h$. 
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 $\alpha_{e1}$ and $\alpha_{e2}$ for $\alpha_1$ and $\alpha_2$, respectively, in the initial state. The calibration of the initial directions of optical axes is to find the values of $\alpha_{e1}$ and $\alpha_{e2}$.

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

$$\tan(\alpha_1 + \alpha_{e1}) + \tan(\alpha_2 + \alpha_{e2}) = \frac{D}{zh}.\quad(22)$$

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

$$a_1xy + a_2x + a_3y + a_4 = 0.\quad(23)$$

where

$$\begin{cases} x = \tan\alpha_{e1} \\ y = \tan\alpha_{e2} \\ a_1 = \tan\alpha_1 + \tan\alpha_2 + \tan\alpha_1 \tan\alpha_2 \frac{D}{zh} - 1 \\ a_2 = \tan\alpha_1 \tan\alpha_2 - \tan\alpha_1 \frac{D}{zh} - 1 \\ a_3 = \tan\alpha_1 \tan\alpha_2 - \tan\alpha_2 \frac{D}{zh} - 1 \\ a_4 = \frac{D}{zh} - \tan\alpha_1 - \tan\alpha_2. \end{cases}\quad(24)$$

Formulas (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 $\alpha_1$ and $\alpha_2$ 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

$$f_i(x, y) = a_1xy + a_2x + a_3y + a_4\quad(25)$$

where $a_1$ to $a_4$ are the coefficients $a_1$ to $a_4$ computed from (24) at the $i$th sampling of calibrating data.

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

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

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...
The image offset coordinates from the image center and the 361 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 $y_{e1}$ and $y_{e2}$ 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 $\alpha_1 = 0$, and $u_{2d}$ and $v_{2d}$ were also generated in the condition $\alpha_2 = 0$.

The coordinates $z_h$ and $x_h$ in frame H were computed using (4) and (5) according to $\alpha_1$ and $\alpha_2$ modified with $\alpha_{c1}$ and $\alpha_{c2}$. With the image coordinates and yawing angles listed in Table I, $y_{e1}$ and $y_{e2}$ were calculated via (9) and (10), except for points 2, 3, 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, $\alpha_1$ and $\alpha_2$ in the numerator of (9) and (10) did not need to be modified with $\alpha_{c1}$ and $\alpha_{c2}$. $\alpha_1$ in the denominator of (9) and $\alpha_2$ in the denominator of (10) were the yawing angles relative to the axis $Z_h$, and they need to be modified with $\alpha_{c1}$ and $\alpha_{c2}$. For points 2, 6, 10, and 14, $y_{e1}$ was calculated via (9) with the image offset coordinates in Table II, $\alpha_1$ in Table I, and $\alpha_{c2}$. For these points via (10) with the image offset coordinates in Table II, $\alpha_2$ in Table I, and $\alpha_{c1}$. The average value of $y_{e1}$ and $y_{e2}$ was taken as the coordinate $y_h$. The experimental results to measure a chessboard were listed in Table III. The data were 385 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 $y_{e1}$ and $y_{e2}$ 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 400

---

**TABLE I**

<table>
<thead>
<tr>
<th>Points</th>
<th>$u_{1d}, v_{1d}$</th>
<th>$u_{2d}, v_{2d}$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-89, 3</td>
<td>91, -112</td>
<td>0.0633</td>
<td>0.1706</td>
</tr>
<tr>
<td>2</td>
<td>1, 4</td>
<td>-2, -112</td>
<td>0.1178</td>
<td>0.1171</td>
</tr>
<tr>
<td>3</td>
<td>89, 4</td>
<td>-91, -113</td>
<td>0.1721</td>
<td>0.0633</td>
</tr>
<tr>
<td>4</td>
<td>179, 5</td>
<td>-181, -114</td>
<td>0.2253</td>
<td>0.0103</td>
</tr>
<tr>
<td>5</td>
<td>-91, 47</td>
<td>95, -68</td>
<td>0.0631</td>
<td>0.1748</td>
</tr>
<tr>
<td>6</td>
<td>-2, 48</td>
<td>3, -68</td>
<td>0.1185</td>
<td>0.1212</td>
</tr>
<tr>
<td>7</td>
<td>88, 47</td>
<td>-92, -68</td>
<td>0.1740</td>
<td>0.0663</td>
</tr>
<tr>
<td>8</td>
<td>179, 49</td>
<td>-181, -68</td>
<td>0.2263</td>
<td>0.0113</td>
</tr>
<tr>
<td>9</td>
<td>-95, 91</td>
<td>98, -23</td>
<td>0.0626</td>
<td>0.1777</td>
</tr>
<tr>
<td>10</td>
<td>-8, 91</td>
<td>5, -24</td>
<td>0.1195</td>
<td>0.1235</td>
</tr>
<tr>
<td>11</td>
<td>87, 91</td>
<td>-90, -22</td>
<td>0.1752</td>
<td>0.0692</td>
</tr>
<tr>
<td>12</td>
<td>178, 91</td>
<td>-181, -81</td>
<td>0.2278</td>
<td>0.0140</td>
</tr>
<tr>
<td>13</td>
<td>-98, 137</td>
<td>100, 22</td>
<td>0.0633</td>
<td>0.1811</td>
</tr>
<tr>
<td>14</td>
<td>-7, 137</td>
<td>8, 24</td>
<td>0.1195</td>
<td>0.1274</td>
</tr>
<tr>
<td>15</td>
<td>85, 136</td>
<td>-88, 25</td>
<td>0.1757</td>
<td>0.0714</td>
</tr>
<tr>
<td>16</td>
<td>176, 136</td>
<td>-181, 25</td>
<td>0.2297</td>
<td>0.0167</td>
</tr>
</tbody>
</table>

**TABLE II**

<table>
<thead>
<tr>
<th>Points</th>
<th>$u_{1d}, v_{1d}$</th>
<th>$u_{2d}, v_{2d}$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>99, 4</td>
<td>-106, -115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>99, 46</td>
<td>-109, -70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>98, 91</td>
<td>-112, -25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>99, 136</td>
<td>-114, 22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE III**

<table>
<thead>
<tr>
<th>Points</th>
<th>X (mm)</th>
<th>Y (mm)</th>
<th>Z (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-6.851</td>
<td>-35.263</td>
<td>559.8564</td>
</tr>
<tr>
<td>2</td>
<td>23.8075</td>
<td>-34.094</td>
<td>556.8797</td>
</tr>
<tr>
<td>3</td>
<td>54.0711</td>
<td>-35.8102</td>
<td>551.6178</td>
</tr>
<tr>
<td>4</td>
<td>83.2076</td>
<td>-34.8537</td>
<td>543.8075</td>
</tr>
<tr>
<td>5</td>
<td>-7.9986</td>
<td>-6.2657</td>
<td>551.5902</td>
</tr>
<tr>
<td>6</td>
<td>22.4331</td>
<td>-6.0727</td>
<td>546.7718</td>
</tr>
<tr>
<td>7</td>
<td>52.8529</td>
<td>-5.5617</td>
<td>541.6156</td>
</tr>
<tr>
<td>8</td>
<td>82.6170</td>
<td>-5.2144</td>
<td>539.8013</td>
</tr>
<tr>
<td>9</td>
<td>-8.8849</td>
<td>22.8316</td>
<td>545.5336</td>
</tr>
<tr>
<td>10</td>
<td>21.8423</td>
<td>22.9313</td>
<td>540.3840</td>
</tr>
<tr>
<td>11</td>
<td>51.6135</td>
<td>23.4384</td>
<td>533.4364</td>
</tr>
<tr>
<td>12</td>
<td>81.0656</td>
<td>23.5013</td>
<td>531.5773</td>
</tr>
<tr>
<td>13</td>
<td>-9.4947</td>
<td>51.5836</td>
<td>538.1701</td>
</tr>
<tr>
<td>14</td>
<td>20.4779</td>
<td>50.8963</td>
<td>532.7708</td>
</tr>
<tr>
<td>15</td>
<td>50.6692</td>
<td>53.0560</td>
<td>528.3036</td>
</tr>
<tr>
<td>16</td>
<td>79.5532</td>
<td>52.1520</td>
<td>522.6241</td>
</tr>
</tbody>
</table>

---

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 $\alpha_1$ and $\alpha_2$ 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 $\times$ 480 in pixel, and its center was [320, 240]. In the experiment, $v_{10}$ and $v_{20}$ 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.
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

<table>
<thead>
<tr>
<th>No</th>
<th>Width (mm)</th>
<th>Height (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.9</td>
<td>31.1</td>
</tr>
<tr>
<td>2</td>
<td>30.6</td>
<td>30.9</td>
</tr>
<tr>
<td>3</td>
<td>30.2</td>
<td>30.8</td>
</tr>
<tr>
<td>4</td>
<td>30.2</td>
<td>30.6</td>
</tr>
<tr>
<td>5</td>
<td>30.5</td>
<td>30.7</td>
</tr>
<tr>
<td>6</td>
<td>30.3</td>
<td>30.5</td>
</tr>
<tr>
<td>7</td>
<td>30.6</td>
<td>30.9</td>
</tr>
<tr>
<td>8</td>
<td>30.9</td>
<td>30.7</td>
</tr>
<tr>
<td>9</td>
<td>30.3</td>
<td>30.5</td>
</tr>
<tr>
<td>10</td>
<td>30.2</td>
<td>30.2</td>
</tr>
<tr>
<td>11</td>
<td>30.5</td>
<td>30.7</td>
</tr>
<tr>
<td>12</td>
<td>30.3</td>
<td>30.5</td>
</tr>
</tbody>
</table>

### Table V

<table>
<thead>
<tr>
<th>Index</th>
<th>X(mm)</th>
<th>Y(mm)</th>
<th>Z(mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-34.1292</td>
<td>16.1916</td>
<td>539.3316</td>
</tr>
<tr>
<td>2</td>
<td>-32.2241</td>
<td>15.7743</td>
<td>693.0157</td>
</tr>
<tr>
<td>3</td>
<td>-49.7623</td>
<td>16.1721</td>
<td>827.8718</td>
</tr>
<tr>
<td>4</td>
<td>-48.2818</td>
<td>5.9594</td>
<td>1065.2969</td>
</tr>
<tr>
<td>5</td>
<td>-67.6102</td>
<td>6.2453</td>
<td>1196.2755</td>
</tr>
<tr>
<td>6</td>
<td>-83.6350</td>
<td>6.8764</td>
<td>1361.9719</td>
</tr>
</tbody>
</table>

### Table VI

<table>
<thead>
<tr>
<th>Index</th>
<th>X(mm)</th>
<th>Y(mm)</th>
<th>Z(mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-34.129</td>
<td>26.1603</td>
<td>539.3316</td>
</tr>
<tr>
<td>2</td>
<td>-32.224</td>
<td>26.2009</td>
<td>693.0157</td>
</tr>
<tr>
<td>3</td>
<td>-49.762</td>
<td>26.5460</td>
<td>827.8718</td>
</tr>
<tr>
<td>4</td>
<td>-48.281</td>
<td>18.3452</td>
<td>1065.2969</td>
</tr>
<tr>
<td>5</td>
<td>-67.610</td>
<td>23.3014</td>
<td>1196.2755</td>
</tr>
<tr>
<td>6</td>
<td>-83.635</td>
<td>26.4331</td>
<td>1361.9719</td>
</tr>
</tbody>
</table>

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

### 403 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: $k_{x1} = 834.82771$, $k_{y1} = 408.815.41740$, $u_{10} = 303.8$, $v_{10} = 306.3$, $k_{x2} = 850.45548$, $k_{y2} = 833.29453$, $u_{20} = 345.1$, 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 $c^{1}T_{c2}$, i.e., the 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 414 positions, i.e.,

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

The experiment scene was similar to that of the initial optical 417 direction calibration, as given in Fig. 8. The intersection be- 418 tween the two black blocks on a target, as shown in Fig. 8, was 419 selected as the point $P$ to be measured. When the target was 420 placed at a position in front of the visual system, the two cam- 421 eros were yawed to initial directions and captured the target’s 422 images. The Cartesian space position of point $P$ in the frame 423
XU et al.: NEW ACTIVE VISUAL SYSTEM FOR HUMANOID ROBOTS

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.

The coordinates of point $P$ in frame $H$ were obtained via transformations including the rotation with $\alpha_{c1}$ 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 $\alpha_1$ and $\alpha_2$, 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 $k_{x1}, k_{y1}, k_{x2}, k_{y2}$, 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 $u_{10} = 320, v_{10} = 240, u_{20} = 437$, 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, 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 of the image of camera $C_{a1}$. In the second step, the cameras were yawed to make the feature point be at the horizontal center of the image of camera $C_{a2}$. $\alpha_1$ and $\alpha_2$ were generated as $\alpha_1 = 0.08 \pm 0.03$ and $\alpha_2 = 0.03 \pm 0.049$. 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 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, and even if the cameras’ intrinsic parameters were not employed, the large distortion in the camera lens was not taken into account in the proposed method.
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 yawn 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


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.
AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1 = The sentence was modified. Is the new sentence appropriate? If not, please provide the necessary corrections.

END OF ALL QUERIES
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.

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.

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: sxu@eudaily.com; tan@compys.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: meyfi@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.
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 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 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 $C_{a2}$. $\Pi$ 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 $\Pi$ with the change in $\theta_2$. Then, it can be on the perpendicular bisector of line $C_1C_2$ on the plane $\Pi$ with the adjustment of $\theta_1$. With simultaneous yawing in opposite directions for the two cameras, the images of point $P$ can be placed at the center 126 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, $\theta_1$ and $\theta_2$ are the joint angles of the two joints.

$$
\begin{bmatrix}
\cos \theta_1 & -\sin \theta_1 & \sin \theta_1 & a_2 \sin \theta_1 \\
\sin \theta_1 & \cos \theta_1 & \cos \theta_1 & -a_2 \cos \theta_1 \\
0 & 0 & 0 & 1
\end{bmatrix}
$$

Assume that the yawing angles of the two cameras are equal to $\alpha_1$. 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 134 axis $Z_h$ is

$$
z_h = D/(2 \tan \alpha_1)
$$

where $D$ is the distance between the optical principal points of the two cameras, and $\alpha_1$ is the yawing angle.

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

$$
\begin{bmatrix}
x_w \\
y_w \\
z_w \\
1
\end{bmatrix}
= wT_h
\begin{bmatrix}
x_h \\
y_h \\
z_h \\
1
\end{bmatrix}
= 
\begin{bmatrix}
-x_h \sin \theta_1 \cos \theta_2 + a_2 \sin \theta_1 \sin \theta_2 \\
z_h \cos \theta_1 \cos \theta_2 - a_2 \cos \theta_1 \sin \theta_2 \\
z_h \sin \theta_2 + a_2 \cos \theta_2 + d_1
\end{bmatrix}
$$

C. General Bionic Visual Model

The general bionic visual model is designed for the robots without the neck. It consists of two cameras simultaneously
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 coordinates of point \( C \) in the second step, the horizontal imaging coordinate of point \( P \) is adjusted to the desired value, and the image coordinates of point \( P \) in camera \( C_{a1} \) 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 \( \alpha_1 \) and \( \alpha_2 \). 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:

\[
\begin{align*}
z_h &= D/(\tan \alpha_1 + \tan \alpha_2) \quad (4) \\
x_h &= z_h \tan \alpha_1 - D/2 \quad (5)
\end{align*}
\]

where \( \alpha_1 \) is the yawing angle in the first step, and \( \alpha_2 \) is the yawing angle in the second step.

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

\[
\begin{align*}
\begin{bmatrix} u_{11} - u_{10} \\ v_{11} - v_{10} \end{bmatrix} &= k_{x1} \begin{bmatrix} x_{c1} \\ y_{c1} \end{bmatrix} \\
\end{align*}
\]

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.,

\[
y_{c1} = \frac{v_{11} - v_{10}}{u_{11} - u_{10}} k_{x1} \approx \frac{v_{1d}}{u_{1d}} x_{c1} \quad (7)
\]

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.,

\[
x_{c1} = \frac{\sin(\alpha_1 - \alpha_2)}{\cos \alpha_1} z_h \quad (8)
\]

where \( \alpha_1 \) and \( \alpha_2 \) are same as described in (4).

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

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

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

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

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 197 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 200
The horizontal coordinate of point $Q_1$ in the first image of camera $C_{a1}$ can be estimated as follows since it is in proportion to the imaging angle:

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

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.,

$$\frac{dN}{h} = \frac{dN}{D} - 2d\alpha_1/\sin(2\alpha_1)$$

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

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

$$\frac{dN}{h} \approx \frac{dN}{D} - d\alpha_1/\alpha_1 \leq |dN/D| + |d\alpha_1/\alpha_1|.$$  

From (16), it is easy to find that the relative error in $h$ is proportional to relative errors $dN/D$ and $d\alpha_1/\alpha_1$. For example, when the relative errors in $D$ and $\alpha_1$ are 1%, the relative error in $h$ 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.,

$$\frac{dN}{h} = \frac{dN}{D} - \frac{(\cos \alpha_2/ \cos \alpha_1) d\alpha_1 + (\cos \alpha_1/ \cos \alpha_2) d\alpha_2}{\sin(\alpha_1 + \alpha_2)}.$$  

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

$$\frac{dN}{h} \approx \frac{dN}{D} - (d(\alpha_1 + \alpha_2)/(\alpha_1 + \alpha_2)$$

$$\leq |dN/D| + |d(\alpha_1 + \alpha_2)/(\alpha_1 + \alpha_2)|.$$  

If $d\alpha_1$ and $d\alpha_2$ are taken as the same, then (17) degenerates 245 to (16).

The term $|d(\alpha_1 + \alpha_2)/(\alpha_1 + \alpha_2)|$ would be large if the 247 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 $Y_h Z_h$ plane. 252 In fact, the influence of the principal point on the errors of $h$ can be taken in the same way as for that of the initial angles and can be reduced via initial angle calibration.

201 the areas $S_3$ and $S_4$, and $S_7$ and $S_8$. The other areas are divided 202 by optical lines $L_{11}$, $L_{12}$, $L_{21}$, and $L_{22}$.

203 Four frames of images are captured at the two measuring 204 positions with yawing angles $\alpha_1$ and $\alpha_2$ for the two cameras.

205 The image coordinates are indicated with $[u_{ijk}, v_{ijk}]$ for object 206 $k$ in the image $j$ of camera $i$. The area in which object $k$ 207 locates can be determined with the image coordinates of object 208 $k$ and the optical principal points, i.e., $[u_{i0j}, v_{i0j}]$, and $[u_{i1j}, v_{i1j}]$, 209 $i = 1, 2$, $j = 1, 2$. The division can be concluded as given in 210 (12) from Fig. 6, i.e.,

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

where $S$ is the area in which the object $k$ locates, $u_{ijk} = u_{ij0} - u_{i0j}$.

After the area in which the object $k$ locates is determined, 214 the approximate position in the area can be estimated according 215 to the image coordinates $u_{ijk}$. In addition, the areas $S_3$ and $S_4$ 216 can be divided into subareas using auxiliary point $Q_1$, which is 217 the intersection of line $B_3 B_4$ and the $Z_h$-axis. The angle $\beta$ is 218 defined as $\angle B_2 C \angle Q_1$, which is given as follows:

$$\beta = \tan(2zh/D) + \alpha_1 - \pi/2.$$
From (9) to (11), the relative error $dy_h/z_h$ is deduced as follows:

$$
\frac{dy_h}{z_h} = \frac{1}{2} \frac{v_{1d}}{u_{1d}} \left[ \frac{\sin(\alpha_1 - \alpha_2)}{\cos \alpha_1} \frac{dz_h}{z_h} + \frac{\cos \alpha_2 \alpha_1 - \cos(\alpha_1 - \alpha_2) \cos \alpha_1 \alpha_2}{(\cos \alpha_1)^2} \right] + \frac{1}{2} \frac{v_{2d}}{u_{2d}} \left[ \frac{\sin(\alpha_1 - \alpha_2)}{\cos \alpha_2} \frac{dz_h}{z_h} + \frac{\cos(\alpha_1 - \alpha_2) \cos \alpha_2 \alpha_1 - \cos \alpha_1 \alpha_2}{(\cos \alpha_2)^2} \right] + \frac{1}{2} \frac{v_{1d} du_{1d}}{u_{1d} \cos \alpha_1} \sin(\alpha_1 - \alpha_2) - \frac{1}{2} \frac{v_{2d} du_{2d}}{u_{2d} \cos \alpha_2} \sin(\alpha_1 - \alpha_2)
$$

(20)

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

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

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

(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 $\alpha_1$ and $\alpha_2$ 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 $y_{x1}$ is calculated with (27), $u_{1d}$ and $v_{1d}$ are generated in the condition $\alpha_2 = 0$. While $y_{x2}$ 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 $\alpha_1 = \pi/6$ and $\alpha_2 = \pi/12$, $u_{1d} = 40$, $v_{1d} = 50$, $u_{2d} = 45$, $v_{2d} = 60$, $dz_h/z_h = 2\%$, $dv_{1d} = 50$, 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 $\alpha_1$ and $\alpha_2$ may be the main source for the relative error in $z_h$.
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 $\alpha_{e1}$ and $\alpha_{e2}$ for $\alpha_1$ and $\alpha_2$, respectively, in the initial state. The calibration of the initial directions of optical axes is to find the values of $\alpha_{e1}$ and $\alpha_{e2}$.

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

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

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

$$a_1xy + a_2x + a_3y + a_4 = 0 \tag{23}$$

where

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

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 $\alpha_1$ and $\alpha_2$ 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 $i$th 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...
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 \( u_{2d} \), and \( y_{1d} \) and \( y_{2d} \) of points 2, 6, 10, and 14 were very small. As analyzed at the end of Section IV, the calculation of \( y_{1c} \) and \( y_{2c} \) would introduce large errors. To deal with this problem, several data were captured for the four points above, that is, \( u_{1d} \) and \( u_{1d} \) in Table I were generated in the condition \( \alpha_1 = 0 \), and \( u_{2d} \) and \( v_{2d} \) were generated in the condition \( \alpha_2 = 0 \).

The coordinates \( z_h \) and \( z_h \) in frame H were computed using (4) and (5) according to \( \alpha_1 \) and \( \alpha_2 \) modified with \( c_{1,1} \) and \( c_{2,1} \). With the image coordinates and yawing angles listed in Table I, \( y_{1c} \) and \( y_{2c} \) 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, \( \alpha_1 \) and \( \alpha_2 \) in the 3D numerators of (9) and (10) did not need to be modified with \( c_{1,1} \), \( c_{1,2} \), \( c_{2,1} \), and \( c_{2,2} \). \( \alpha_1 \) in the denominator of (9) and (10) were the yawing angles relative to the axis \( Z_h \), and they were 380.

For points 2, 6, 10, and 14, \( y_{1c} \) was calculated via (9) (with the image offset coordinates \( c_{1,1} \) and \( c_{1,2} \)) for points via (10) with the image offset coordinates \( c_{2,1} \) and \( c_{2,2} \). The average value of \( y_{1c} \) and \( y_{2c} \) was taken as the coordinate \( y_{1c} \). The experimental results to 384 measure a chessboard were listed in Table III. The data were 385 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 widths and heights computed from the distances between any two 390 adjacent cross points in the pattern were listed in Table IV and were 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 395 the difference of \( y_{1c} \) and \( y_{2c} \) 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 399 the measuring accuracy with the proposed visual system and 400 the two cameras might not be parallel.

### Table I

<table>
<thead>
<tr>
<th>Points</th>
<th>( u_{1d} )</th>
<th>( v_{1d} )</th>
<th>( u_{2d} )</th>
<th>( v_{2d} )</th>
<th>( \alpha_1 )</th>
<th>( \alpha_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-89, 3</td>
<td>91, 112</td>
<td>0.0633</td>
<td>0.1706</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1, 4</td>
<td>-2, 112</td>
<td>0.1178</td>
<td>0.1117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>89, 4</td>
<td>-91, 113</td>
<td>0.1721</td>
<td>0.0633</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>179, 5</td>
<td>-181, 114</td>
<td>0.2253</td>
<td>0.0103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-91, 47</td>
<td>95, 68</td>
<td>0.0631</td>
<td>0.1748</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-2, 48</td>
<td>3, 68</td>
<td>0.1185</td>
<td>0.1212</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>88, 47</td>
<td>-92, 68</td>
<td>0.1740</td>
<td>0.0663</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>179, 49</td>
<td>-181, 68</td>
<td>0.2263</td>
<td>0.0113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-95, 91</td>
<td>98, 23</td>
<td>0.0626</td>
<td>0.1777</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-4, 91</td>
<td>5, 24</td>
<td>0.1195</td>
<td>0.1235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>87, 91</td>
<td>-90, 22</td>
<td>0.1752</td>
<td>0.0692</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>178, 91</td>
<td>-181, 21</td>
<td>0.2278</td>
<td>0.0140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>-98, 137</td>
<td>100, 22</td>
<td>0.0633</td>
<td>0.1811</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>-7, 137</td>
<td>8, 24</td>
<td>0.1195</td>
<td>0.1274</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>85, 136</td>
<td>-88, 25</td>
<td>0.1757</td>
<td>0.0714</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>176, 136</td>
<td>-181, 25</td>
<td>0.2297</td>
<td>0.0167</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th>Points</th>
<th>( u_{1d} )</th>
<th>( v_{1d} )</th>
<th>( u_{2d} )</th>
<th>( v_{2d} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>99, 4</td>
<td>-106, -115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>99, 46</td>
<td>-109, -70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>98, 91</td>
<td>-112, -25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>99, 136</td>
<td>-114, 22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 \( c_1T_{c2} \), i.e., the 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.,

\[
c_1T_{c2} = \begin{bmatrix}
0.9995 & -0.0236 & -0.0222 & -150.9556 \\
0.0234 & 0.9997 & -0.0091 & -5.1851 \\
0.0224 & 0.0086 & 0.9997 & 2.2226 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]
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.

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, 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 $C_{a2}$. $\alpha_1$ and $\alpha_2$ 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 calculated in the case that the intrinsic parameters of the cameras were supposed to be not available.

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, and the large distortion in the camera lens was not taken into account in the proposed method.
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


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.
AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1 = The sentence was modified. Is the new sentence appropriate? If not, please provide the necessary corrections.

END OF ALL QUERIES