23

Visual Guided Approach-to-grasp for Humanoid Robots

Yang Shen[1], De Xu[1], Min Tan[1] and Ze-Min Jiang[2]
[1]*Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences*
[2]*School of Information Science and Technology, Beijing Institute of Technology*
P. R. China

1. Introduction

Vision based control for robots has been an active area of research for more than 30 years and significant progresses in the theory and application have been reported (Hutchinson et al., 1996; Kragic & Christensen, 2002; Chaumette & Hutchinson, 2006). Vision is a very important non-contact measurement method for robots. Especially in the field of humanoid robots, where the robot works in an unstructured and complex environment designed for human, visual control can make the robot more robust and flexible to unknown changes in the environment (Hauck et al., 1999).

Humanoid robot equipped with vision system is a typical hand-eye coordination system. With cameras mounted on the head, the humanoid robot can manipulate objects with his hands. Generally, the most common task for the humanoid robot is the approach-to-grasp task (Horaud et al., 1998). There are many aspects concerned with the visual guidance of a humanoid robot, such as vision system configuration and calibration, visual measurement, and visual control.

One of the important issues in applying vision system is the calibration of the system, including camera calibration and head-eye calibration. Calibration has received wide attentions in the communities of photogrammetry, computer vision, and robotics (Clarke & Fryer, 1998). Many researchers have contributed elegant solutions to this classical problem, such as Faugeras and Toscani, Tsai, Heikkila and Silven, Zhang, Ma, Xu. (Faugeras & Toscani, 1986; Tsai, 1987; Heikkila & Silven, 1997; Zhang, 2000; Ma, 1996; Xu et al., 2006a). Extensive efforts have been made to achieve the automatic or self calibration of the whole vision system with high accuracy (Tsai & Lenz, 1989). Usually, in order to gain a wide field of view, the humanoid robot employs cameras with lens of short focal length, which have a relatively large distortion. This requires a more complex nonlinear model to represent the distortion and makes the accurate calibration more difficult (Ma et al., 2003).

Another difficulty in applying vision system is the estimation of the position and orientation of an object relative to the camera, known as visual measurement. Traditionally, the position of a point can be determined with its projections on two or more cameras based on epipolar geometry (Harley & Zisserman, 2004). Han et al. measured the pose of a door knob relative to the end-effector of the manipulator with a specially designed mark attached on the knob

Source: Humanoid Robots: Human-like Machines, Book edited by: Matthias Hackel
ISBN 978-3-902613-07-3, pp. 642, Itech, Vienna, Austria, June 2007

(Han et al., 2002). Lack of constraints, errors in calibration and noises on feature extraction restrict the accuracy of the measurement. When the structure or the model of the object is prior known, it can be taken to estimate the pose of the object by means of matching. Kragic et al. taken this technique to determine the pose of the workpiece based on its CAD model (Kragic et al., 2001). High accuracy can be obtained with this method for the object of complex shape. But the computational consumption needed for matching prevents its application from real-time measurement. Therefore, accuracy, robustness and performance are still the challenges for visual measurement.

Finally visual control method also plays an important role in the visual guided approach-to-grasp movement of the humanoid robot. Visual control system can be classified into eye-to-hand (ETH) system and eye-in-hand (EIH) system based on the employed camera-robot configuration (Hutchinson et al., 1996). An eye-to-hand system can have a wider field of view since the camera is fixed in the workspace. Hager et al. presented an ETH stereo vision system to position two floppy disks with the accuracy of 2.5mm (Hager et al., 1995). Hauck et al. proposed a system for grasping (Hauck et al., 2000). On the other hand, an eye-in-hand system can possess a higher precision as the camera is mounted on the end-effector of the manipulator and can observe the object more closely. Hashimoto et al. (Hashimoto et al., 1991) gave an EIH system for tracking. According to the ways of using visual information, visual control can also be divided into position-based visual servoing (PBVS), image-based visual servoing (IBVS) and hybrid visual servoing (Hutchinson et al., 1996; Malis et al., 1999; Corke & Hutchinson, 2001). Dodds et al. pointed out that a key to solving robotic hand-eye tasks efficiently and robustly is to identify how precise the control is needed at a particular time during task execution (Dodds et al., 1999). With the hierarchical architecture he proposed, a hand-eye task was decomposed into a sequence of primitive sub tasks. Each sub task had a specific requirement. Various visual control techniques were integrated to achieve the whole task. A similar idea was demonstrated by Kragic and Christensen (Kragic & Christensen, 2003). Flandin et al. combined ETH and EIH together to exploit the advantage of both configurations (Flandin et al., 2000). Hauck et al. integrated look-and-move with position-based visual servoing to achieve 3 degrees of freedom (DOFs) reaching task (Hauck et al., 1999).

In this chapter, issues above are discussed in detail. Firstly, a motion based method is provided to calibrate the head-eye geometry. Secondly, a visual measurement method with shape constraint is presented to determine the pose of a rectangle object. Thirdly, a visual guidance strategy is developed for the approach-to-grasp movement of humanoid robots.

The rest of the chapter is organized as follows. The camera-robot configuration and the assignment of the coordinate frames for the robot are introduced in section 2. The calibration of vision system is investigated in section 3. In this section, the model for cameras with distortion is presented, and the position and orientation of the stereo rig relative to the head can be determined with three motions of the robot head. In section 4, the shape of a rectangle is taken as the constraint to estimate the pose of the object with high accuracy. In section 5, the approach-to-grasp movement of the humanoid robot is divided into five stages, namely searching, approaching, coarse alignment, precise alignment and grasping. Different visual control methods, such as ETH/EIH, PBVS/IBVS, look-then-move/visual servoing, are integrated to accomplish the grasping task. An experiment of valve operation by a humanoid robot is also presented in this section. The chapter is concluded in section 6.

2. Camera-robot configuration and robot frame

A humanoid robot[1] has the typical configuration of vision system as shown in Fig. 1 (Xu et al., 2006b). Two cameras are mounted on the head of the robot, which serve as eyes. The arms of the robot serve as manipulators with grippers attached at the wrist as the hands. An eye-to-hand system is formed with these two cameras and the arms of the robot. If another camera is mounted on the wrist, an eye-in-hand system will be formed.

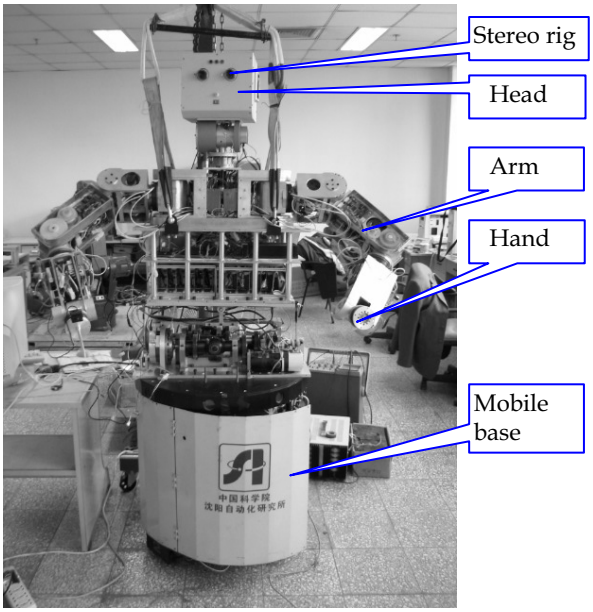


Figure 1. Typical configuration of humanoid robots

Throughout this chapter, lowercase letters (a, b, c) are used to denote scalars, bold-faced ones (**a**, **b**, **c**) denote vectors. Bold-faced uppercase letters (**A, B**, **C**) stand for matrices and italicized uppercase letters (*A*, *B*, *C*) denote coordinate frames. The homogeneous transformation from coordinate frame *X* to frame *Y* is denoted by ${}^{y}\mathbf{T}_{x}$. It is defined as follows:

$$ {}^{y}T_{x} = \begin{bmatrix} {}^{y}R_{x} & {}^{y}p_{x0} \\ 0 & 1 \end{bmatrix} \tag{1} $$

where ${}^{y}\mathbf{R}_{x}$ is a 3 x 3 rotation matrix, and ${}^{y}\mathbf{p}_{x0}$ is a 3 x 1 translation vector.
Figure 2 demonstrates the coordinate frames assigned for the humanoid robot. The subscript *B*, *N*, *H*, *C*, *G* and *E* represent the base frame of the robot, the neck frame, the head frame, the camera frame, the hand frame, and the target frame respectively. For example, ${}^{n}\mathrm{T}_{h}$ represents the pose (position and orientation) of the head relative to the neck.

[1] The robot is developed by Shenyang Institute of Automation, cooperated with Institute of Automation, Chinese Academy of Sciences, P. R. China.

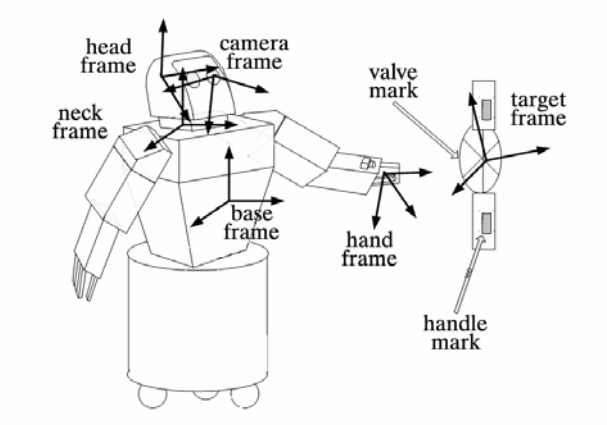


Figure 2. Coordinate frames for the robot

The head has two DOFs such as yawing and pitching. The sketch of the neck and head of a humanoid robot is given in Fig. 3. The first joint is responsible for yawing, and the second one for pitching. The neck frame *N* for the head is assigned at the connection point of the neck and body. The head frame *H* is assigned at the midpoint of the two cameras. The coordinate frame of the stereo rig is set at the optical center of one of the two cameras, e.g. the left camera as shown in Fig. 3.

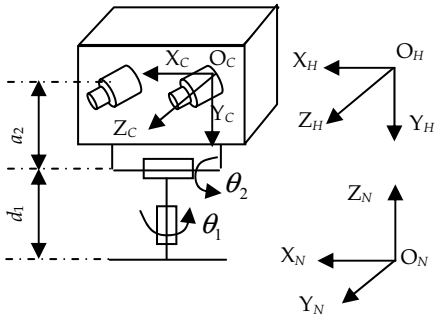


Figure 3. Sketch of the neck and the head

From Fig. 3, the transformation matrix from the head frame *H* to the neck frame *N* is given in (2) according to Denavit-Hartenberg (D-H) parameters (Murray et al., 1993).

$$ {}^{n}T_{c} = \begin{bmatrix} c\theta_1 & -s\theta_1 s\theta_2 & -s\theta_1 c\theta_2 & a_2 s\theta_1 s\theta_2 \\ s\theta_1 & c\theta_1 s\theta_2 & c\theta_1 c\theta_2 & -a_2 c\theta_1 s\theta_2 \\ 0 & -c\theta_2 & s\theta_2 & a_2 c\theta_2 + d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2} $$

where d_1 and a_2 are the D-H parameters for the two links of the head, θ_1 and θ_2 are the corresponding joint angles, $c\theta$ denotes $\cos\theta$, and $s\theta$ denotes $\sin\theta$.

3. Vision system calibration

3.1 Camera model

The real position of a point on the image plane will deviate from its ideal position as a result of the distortion of the optical lens components. Let (u, v) denote the real pixel coordinates for the projection of a point, (u′, v′) denote the ideal pixel coordinates without distortion. The distortion is defined as follows:

$$\begin{cases} u = u' + \delta_u(u', v') \\ v = v' + \delta_v(u', v') \end{cases} \tag{3}$$

where δ_u and δ_v represent the distortion in the horizontal and vertical directions respectively.

The distortion can be modeled as a high order polynomial which contains both radial and tangential distortion (Ma et al., 2003). Generally the distortion is formed mainly by the radial component, so the following second order radial distortion model without tangential component is employed for cameras with standard field of view:

$$\begin{cases} u - u_0 = (u' - u_0)(1 + k_u' r'^2) \\ v - v_0 = (v' - v_0)(1 + k_v' r'^2) \end{cases} \tag{4}$$

where (u_0, v_0) are the pixel coordinates of the principle point, (k_u', k_v') are the radial distortion coefficients, and $r' = \sqrt{(u' - u_0)^2 + (v' - v_0)^2}$ is the radius from the ideal point (u′, v′) to the principle point (u_0, v_0).

When correcting the distortion, the distorted image needs to be corrected to a linear one. So the reverse problem of (4) needs to be solved to obtain the ideal pixel coordinates (u′, v′) from (u, v). Then the following model is adopted instead of (4):

$$\begin{cases} u'' - u_0 = (u - u_0)(1 + k_u r^2) \\ v'' - v_0 = (v - v_0)(1 + k_v r^2) \end{cases} \tag{5}$$

where (u′′, v′′) are the pixel coordinates after distortion correlation, k_u, k_v are the correction coefficients, $r = \sqrt{(u - u_0)^2 + (v - v_0)^2}$ is the radius from the point (u, v) to the principle point.

After applying distortion correction, the pixel coordinates (u′′, v′′) for the projection of a point in the camera frame can be determined with the intrinsic parameters of the camera. Here the four parameters model, which does not consider the skew between the coordinate axes, is employed as follows:

$$\begin{bmatrix} u'' \\ v'' \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & u_0 \\ 0 & k_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c / z_c \\ y_c / z_c \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} x_c / z_c \\ y_c / z_c \\ 1 \end{bmatrix} \tag{6}$$

where (x_c, y_c, z_c) are the coordinates of a point in the camera frame, (k_x, k_y) are the focal length in pixel, $\mathbf{M_1}$ is known as the intrinsic parameter matrix of the camera.

Assume the coordinates of a point in the world reference frame *W* is (x_w, y_w, z_w). Let (x_c, y_c, z_c) be the coordinates of the point in the camera reference frame. Then (x_w, y_w, z_w) and (x_c, y_c, z_c) are related to each other through the following linear equation:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_2 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{7}$$

where $\mathbf{n} = (n_x, n_y, n_z)^T$, $\mathbf{o} = (o_x, o_y, o_z)^T$ and $\mathbf{a} = (a_x, a_y, a_z)^T$ are the coordinate vectors for the x-axis, y-axis and z-axis of the world frame in the camera frame, $\mathbf{p} = (p_x, p_y, p_z)^T$ is the coordinate vector of the origin for the world reference frame in the camera frame, $\mathbf{M_2}$ is a 3 x 4 matrix, which is known as the extrinsic parameter matrix of the camera.

3.2 Hand-eye calibration

For a stereo rig, the intrinsic parameters of each camera and the displacement between two cameras can be determined accurately with the method proposed by Xu et al., which is designed for cameras with large lens distortion (Xu et al., 2006a). Then the position of a point in the camera reference frame can be determined with this calibrated stereo rig. The next important step in applying the stereo rig on the humanoid robot is to determine the relative position and orientation between the stereo rig and the head of the robot, which is called head-eye (or hand-eye) calibration.

3.2.1 Calibration algorithm

Refer to Fig. 2, assume that the world coordinate frame is attached on the grid pattern (called calibration reference). The pose of the world frame relative to the camera can be determined with the stereo rig by using the grid pattern. If $\mathbf{T_c}$ represents the transformation from the world reference frame to the camera frame; $\mathbf{T_h}$ is the relative pose of the head with respect to the base of the humanoid robot; $\mathbf{T_m}$ represents the head-eye geometry, which is the pose of the stereo rig relative to the robot head. Then it can be obtained that

$$T_p = T_h T_m T_c \tag{8}$$

where $\mathbf{T}_p$ is the transformation between the grid pattern and the robot base.
With the position and orientation of the grid pattern fixed while the pose of the head varying, it can be obtained that

$$T_{hi} T_m T_{ci} = T_{hi-1} T_m T_{ci-1} \tag{9}$$

where the subscript *i* represents the *i*-th motion, $i = 1, 2, \ldots, n$, $\mathbf{T_{h0}}$ and $\mathbf{T_{c0}}$ represent the initial position of the robot head and the camera.
Left multiplying both sides of (9) by $\mathbf{T_{hi\text{-}1}}^{-1}$ and right multiplying by $\mathbf{T_{ci}}^{-1}$ gives:

$$T_{hi-1}^{-1} T_{hi} T_m = T_m T_{ci-1} T_{ci}^{-1} \tag{10}$$

Let $T_{Li} = T_{hi-1}^{-1} T_{hi}$ and $T_{Ri} = T_{ci-1} T_{ci}^{-1}$. $\mathbf{T_{Li}}$ is the transformation between the head reference frames before and after the motion, which can be read from the robot controller. And $\mathbf{T_{Ri}}$ is the transformation between the camera reference frames before and after the movement, which

can be determined by means of stereovision method using the grid pattern. Then (10) becomes:

$$T_{Li}T_m = T_mT_{Ri} \tag{11}$$

Solving (11) will give the head-eye geometry $\mathbf{T_m}$. Equation (11) is the basic equation of head-eye calibration, which is called the **AX = XB** equation in the literature. Substituting (1) into (11) gives:

$$\begin{bmatrix} R_{Li}R_m & R_{Li}p_m + p_{Li} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R_mR_{Ri} & R_mp_{Ri} + p_m \\ 0 & 1 \end{bmatrix} \tag{12}$$

where $\mathbf{R_{Li}}$, $\mathrm{R_{Ri}}$ and $\mathbf{R_m}$ are the rotation components of $\mathbf{T_{Li}}$, $\mathbf{T_{Ri}}$ and $\mathbf{T_m}$ respectively, $\mathbf{p_{Li}}$, $\mathbf{p_{Ri}}$ and $\mathbf{p_m}$ are the translation components of $\mathbf{T_{Li}}$, $\mathbf{T_{Ri}}$ and $\mathbf{T_m}$. By (12) it can be obtained that

$$R_{Li}R_m = R_mR_{Ri} \tag{13}$$

$$R_{Li}p_m + p_{Li} = R_mp_{Ri} + p_m \tag{14}$$

Then $\mathbf{R_m}$ and $\mathbf{p_m}$ can be determined with (13) and (14).

3.2.2 Calibration of the rotation component $\mathbf{R_m}$

A rotation **R** can be represented as (Murray et al., 1993):

$$\mathbf{R} = \mathrm{Rot}(\boldsymbol{\omega}, \theta) \tag{15}$$

where Rot(.) is a function representing the rotation about an axis with an angle, $\boldsymbol{\omega}$ is a unit vector which specific the axis of rotation, θ is the angle of rotation. $\boldsymbol{\omega}$ and θ can be uniquely determined from **R** (Murray et al., 1993). The vector $\boldsymbol{\omega}$ is also the only real eigenvector of **R** and its corresponding eigenvalue is 1 (Tsai & Lenz, 1989):

$$\mathbf{R}\boldsymbol{\omega} = \boldsymbol{\omega} \tag{16}$$

Applying (16) to the $\mathbf{R_{Li}}$ and $\mathbf{R_{Ri}}$ in (13) gives (Tsai & Lenz, 1989):

$$\boldsymbol{\omega}_{\mathrm{Li}} = \mathbf{R}_{\mathrm{m}}\, \boldsymbol{\omega}_{\mathrm{Ri}} \tag{17}$$

where $\boldsymbol{\omega}_{\mathrm{Li}}$ and $\boldsymbol{\omega}_{\mathrm{Ri}}$ are the rotation axes of $\mathbf{R_{Li}}$ and $\mathbf{R_{Ri}}$ respectively.

Let $\omega_{Li} = (\omega_{Lix}, \omega_{Liy}, \omega_{Liz})^T$, $\omega_{Ri} = (\omega_{Rix}, \omega_{Riy}, \omega_{Riz})^T$ and $R_m = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & m_9 \end{bmatrix}$.

Then (17) becomes:

$$A_{mi}x_C = b_{mi} \tag{18}$$

where

$$A_{mi} = \begin{bmatrix} \omega_{Rix} & \omega_{Riy} & \omega_{Riz} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_{Rix} & \omega_{Riy} & \omega_{Riz} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \omega_{Rix} & \omega_{Riy} & \omega_{Riz} \end{bmatrix},$$

$b_{mi} = \omega_{Li} = (\omega_{Lix}, \omega_{Liy}, \omega_{Liz})^T$, $x_C = \begin{pmatrix} m_1 & m_2 & \cdots & m_9 \end{pmatrix}^T$ which is a 9 x 1 vector formed with the columns of the rotation matrix **R**.

Stacking (18) gives the result for the case of n movements of the robot head. A linear equation can be obtained:

$$A_m x_C = b_m \tag{19}$$

where $A_m = \begin{bmatrix} A_{m1} \\ A_{m2} \\ \vdots \\ A_{mn} \end{bmatrix}$ is a $3n$ x 9 matrix, $b_m = \begin{bmatrix} b_{m1} \\ b_{m2} \\ \vdots \\ b_{mn} \end{bmatrix}$ is $3n$ x 1 vector.

Equation (18) indicates that one motion of the head will contribute three equations in (19). Therefore three motions are necessary in order to determine $\mathbf{x_C}$ which has nine independent variables. Providing $n \geq 3$, a least square solution for $\mathbf{x_C}$ is given by

$$x_C = (A_m^T A_m)^{-1} A_m^T b_m \tag{20}$$

Then the rotation component $\mathbf{R_m}$ can be determined from $\mathbf{x_C}$.

3.2.3 Orthogonalization of $\mathbf{R_m}$

The result from previous section gives an estimation of $\mathbf{R_m}$. The deduction of (20) does not consider the orthogonality of $\mathbf{R_m}$. It is necessary to orthogonalize the $\mathbf{R_m}$ obtained from (20).

Assume α, β, γ are the Euler angles of the rotation. Then $\mathbf{R_m}$ can be represented as follows (Murray et al., 1993):

$$R_m(\alpha, \beta, \lambda) = Rot(z, \alpha) Rot(y, \beta) Rot(z, \gamma) \tag{21}$$

where Rot(y, .) and Rot(z, .) are functions representing the rotation about y-axis and z-axis.

Equation (21) yields that $\mathbf{R_m}$ is a nonlinear function of α, β and γ. Then the problem of solving (17) can be formulated as a nonlinear least squares optimization. The objective function to be minimized, J, is a function of the squared error:

$$J(\alpha, \beta, \lambda) = \sum_{i=1}^{n} \left\| \omega_{Li} - R_m(\alpha, \beta, \lambda) \omega_{Ri} \right\|^2 \tag{22}$$

The objective function can be minimized using a standard nonlinear optimization method such as Quasi-Newton method.

$$(\alpha^*, \beta^*, \gamma^*) = \min_{\alpha, \beta, \lambda} J(\alpha, \beta, \lambda) \tag{23}$$

where α^*, β^*, γ^* are the angles where the objective function J reaches its local minimum.

Finally the $\mathbf{R_m}$ is determined by substituting α^*, β^* and γ^* into (21). The orthogonality of $\mathbf{R_m}$ is satisfied since the rotation is represented with the Euler angle as (21). The result from (20) can be taken as the initial value to start the iteration of the optimization method.

3.2.4 Calibration of the translation component $\mathbf{p_m}$

The translation vector $\mathbf{p}_m$ can be determined from (14) once the R_m has been obtained. Rearranging (14) gives:

$$(R_{Li} - I)p_m = R_m p_{Ri} - p_{Li} \tag{24}$$

where **I** stands for a 3 x 3 identity matrix.
It is similar to the derivation from (18) to (19), that a linear equation can be formulated by stacking (24) with the subscript i increasing from 1 to n:

$$C_m p_m = d_m \tag{25}$$

Where $C_m = \begin{bmatrix} R_{L1} - I \\ R_{L2} - I \\ \vdots \\ R_{Ln} - I \end{bmatrix}$ is a $3n$ x 3 matrix, $d_m = \begin{bmatrix} R_m p_{R1} - p_{L1} \\ R_m p_{R2} - p_{L2} \\ \vdots \\ R_m p_{Rn} - p_{Ln} \end{bmatrix}$ is a $3n$ x 1 vector.

Solving (25) gives the translation component $\mathbf{p_m}$ of the head-eye geometry. Giving n ≥ 1, the least square solution for (25) is as follows:

$$p_m = (C_m^T C_m)^{-1} C_m^T d_m \tag{26}$$

3.3 Experiments and results

The head was fixed at the end of a K-10 manipulator as shown in Fig. 4. A stereo rig was mounted on the head and was faced to the ground. A grid pattern was placed under the head. The world reference frame was attached on the grid pattern with its origin at the center of the pattern. The reference frame of the stereo rig was assigned to the frame of the left camera. The stereo rig was calibrated with the method in (Xu et al., 2006a). The intrinsic parameters of each camera of the stereo rig are shown in Table 1.

Figure 4. Head-eye calibration

	k_u (x 10^{-7})	k_v (x 10^{-7})	u_0	v_0	k_x	k_y
Left	4.2180	3.6959	324.6	248.3	1082.3	1067.8
Right	3.4849	3.6927	335.0	292.0	1252.2	1242.3

Table 1. Parameters of the stereo rig

$\mathbf{T_{ci}}$	$\mathbf{T_{hi}}$	$\mathbf{T_{pi}}$
0.0162 -0.1186 0.9928 1100.3 0.9269 -0.3706 -0.0594 -275.3 0.3749 0.9212 0.1040 357.7	0.9989 0.0195 -0.0418 2.0 -0.0362 0.8924 -0.4497 -9.0 0.0286 0.4508 0.8922 320.4	0.9989 −0.0437 0.0152 1148.50 −0.0439 −0.9990 0.0092 −323.55 0.0148 −0.0098 −0.9998 0.02
0.9989 0.0195 -0.0418 2.0 -0.0362 0.8924 -0.4497 -9.0 0.0286 0.4508 0.8922 320.4	0.9675 -0.1499 0.2037 -1.0 0.2261 0.8738 -0.4306 -10.1 -0.1134 0.4627 0.8792 313.2	0.9992 −0.0397 0.0082 1148.60 −0.0398 −0.9991 0.0154 −322.55 0.0076 −0.0158 −0.9998 −0.92
0.9675 -0.1499 0.2037 -1.0 0.2261 0.8738 -0.4306 -10.1 -0.1134 0.4627 0.8792 313.2	0.9675 -0.1499 0.2037 -1.0 0.2261 0.8738 -0.4306 -10.1 -0.1134 0.4627 0.8792 313.2	0.9994 −0.0349 0.0085 1150.20 −0.0350 −0.9994 0.0045 −325.44 0.0083 −0.0049 −0.9999 −0.22
0.9675 -0.1499 0.2037 -1.0 0.2261 0.8738 -0.4306 -10.1 -0.1134 0.4627 0.8792 313.2	0.9675 -0.1499 0.2037 -1.0 0.2261 0.8738 -0.4306 -10.1 -0.1134 0.4627 0.8792 313.2	0.9988 −0.0489 0.0141 1150.10 −0.0489 −0.9988 0.0069 −324.56 0.0137 −0.0075 −0.9999 −0.66

Table 2. The obtained $\mathbf{T_{ci}}$, $\mathbf{T_{hi}}$, and $\mathbf{T_{pi}}$

The displacement between two cameras, which is denoted by ${}^{R}\mathbf{T_{L}}$, is as follows:

$$ {}^{R}T_{L} = \begin{bmatrix} 0.9998 & -0.0198 & -0.0063 & -93.4482 \\ 0.0195 & 0.9990 & -0.0403 & 1.4111 \\ 0.0071 & 0.0402 & 0.9991 & 125.7946 \end{bmatrix}. $$

Four pairs of images were acquired by the stereo rig with 3 motions of the head. The relative position and orientation of the grid pattern with respect to the stereo rig, $\mathbf{T_{ci}}$, was measured with the stereovision method. The pose of the head was changed by the movement of the manipulator, while holding the grid pattern in the field of view of the stereo rig. The pose of the end of the manipulator, $\mathbf{T_{hi}}$, was read form the robot controller. Then 3 equations as (11) were obtained. The head-eye geometry, $\mathbf{T_{m}}$, was computed with (20), (21), (23) and (26). The obtained $\mathbf{T_{ci}}$ and $\mathbf{T_{hi}}$ are shown in the first two columns of Table 2, and the calibration result is as follows:

$$ T_{m} = \begin{bmatrix} -0.0217 & -0.6649 & -0.7466 & 55.1906 \\ -0.0426 & 0.7467 & -0.6638 & -97.8522 \\ 0.9989 & 0.0174 & -0.0446 & 26.0224 \end{bmatrix}. $$

The pose of the grid pattern relative to the robot could be determined by (8) with each group of $\mathbf{T_{ci}}$, $\mathbf{T_{hi}}$ and the obtained $\mathbf{T_{m}}$, as shown in the last column of Table 2. Since the pattern was fixed during the calibration, $\mathbf{T_{pi}}$ should remain constant. From Table 2, the maximum variances of the x, y, z coordinates of the translation vector in $\mathbf{T_{pi}}$ were less than 1.7mm, 2.9mm, and 1.0mm. The results indicated that the head-eye calibration was accurate.

4. Stereo vision measurement for humanoid robots

4.1 Visual positioning with shape constraint

The position and orientation of a object relative to the robot can be measured with the stereo rig on the robot head after the vision system and the head-eye geometry are calibrated. Generally, it is hard to obtain high accuracy with visual measurement, especially the measurement of the orientation, only using individual feature points. Moreover, errors in calibration and feature extraction result in large errors in pose estimation. The estimation performance is expected to be improved if the shape of an object is taken into account in the visual measurement.

Rectangle is a category of shape commonly encountered in everyday life. In this section, the shape of a rectangle is employed as a constraint for visual measurement. A reference frame is attached on the rectangle as shown in Fig. 5. The plane containing the rectangle is taken as the xoy plane. Then the pose of the rectangle with respect to the camera is exactly the extrinsic parameters of the camera if the reference frame on the rectangle is taken as the world reference frame. Assume the rectangle is $2X_w$ in width and $2Y_w$ in height. Obviously, any point on the rectangle plane should satisfy $z_w = 0$.

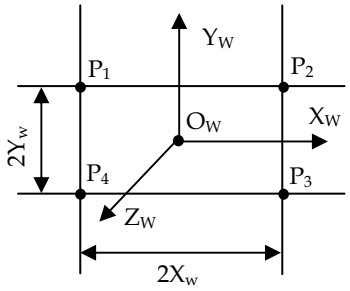


Figure 5. The reference frame and the reference point

4.2 Algorithm for estimating the pose of a rectangle

4.2.1 Derivation of the coordinate vector of x-axis

From (7), and according to the orthogonality of the rotation component of the extrinsic parameter matrix $\mathbf{M_2}$, giving $z_w = 0$, it can be obtained that

$$\begin{cases} o_x x_c + o_y y_c + o_z z_c = y_w + o_x p_x + o_y p_y + o_z p_z \\ a_x x_c + a_y y_c + a_z z_c = a_x p_x + a_y p_y + a_z p_z \end{cases} \tag{27}$$

Assuming that $z_c \neq 0$, $a_x p_x + a_y p_y + a_z p_z \neq 0$, equation (27) becomes:

$$\frac{o_x x'_c + o_y y'_c + o_z}{a_x x'_c + a_y y'_c + a_z} = C_1 \tag{28}$$

where $x'_c = x_c / z_c$, $y'_c = y_c / z_c$ and

$$C_1 = \frac{y_w + o_x p_x + o_y p_y + o_z p_z}{a_x p_x + a_y p_y + a_z p_z}.$$

Points on a line paralleled to the x-axis have the same y coordinate y_w, so C_1 is constant for these points. Taking two points on this line, denoting their coordinates in the camera frame as (x_{ci}, y_{ci}, z_{ci}) and (x_{cj}, y_{cj}, z_{cj}), and applying them to (28) gives:

$$\frac{o_x x'_{ci} + o_y y'_{ci} + o_z}{a_x x'_{ci} + a_y y'_{ci} + a_z} = \frac{o_x x'_{cj} + o_y y'_{cj} + o_z}{a_x x'_{cj} + a_y y'_{cj} + a_z} \tag{29}$$

Simplifying (29) with the orthogonality of the rotation components of $\mathbf{M_2}$ gives:

$$n_x(y'_{ci} - y'_{cj}) + n_y(x'_{cj} - x'_{ci}) + n_z(x'_{ci} y'_{cj} - x'_{cj} y'_{ci}) = 0 \tag{30}$$

Noting that x'_{ci}, y'_{ci}, x'_{cj} and y'_{cj} can be obtained with (5) and (6) if the parameters of the camera have been calibrated, n_x, n_y and n_z are the only unknowns in (30). Two equations as (30) can be obtained with two lines paralleled to the x-axis, and besides, **n** is a unit vector, i.e. $\|\mathbf{n}\| = 1$. Then n_x, n_y and n_z can be determined with these three equations.
It can be divided into two cases to obtain n_x, n_y and n_z. If the optical axis of the camera is not vertical to the rectangle plane, $n_z \neq 0$ is satisfied. Dividing both sides of (30) by n_z gives:

$$n'_x(y'_{ci} - y'_{cj}) + n'_y(x'_{cj} - x'_{ci}) = x'_{cj}y'_{ci} - x'_{ci}y'_{cj} \tag{31}$$

where $n'_x = n_x/n_z$, $n'_y = n_y/n_z$.
Then n'_x and n'_y can be determined with two such equations. The corresponding n_x, n_y, n_z can be computed by normalizing the vector $(n'_x, n'_y, 1)^T$ as follows:

$$\begin{cases} n_x = n'_x \Big/ \sqrt{n'^2_x + n'^2_y + 1} \\ n_y = n'_y \Big/ \sqrt{n'^2_x + n'^2_y + 1} \\ n_z = 1 \Big/ \sqrt{n'^2_x + n'^2_y + 1} \end{cases} \tag{32}$$

If the optical axis is vertical to the rectangle plane, $n_z = 0$ and (30) becomes:

$$n_x(y'_{ci} - y'_{cj}) + n_y(x'_{cj} - x'_{ci}) = 0 \tag{33}$$

Similar to (31), n_x and n_y can be directly computed with two equations as (33), and the n_x, n_y, n_z can be obtained by normalizing the vector $(n_x, n_y, 0)^T$ to satisfy $\|\mathbf{n}\| = 1$.

4.2.2 Derivation of the coordinate vector of z-axis

Similar to (27), by $\mathbf{M_2}$, it can be obtained that:

$$\begin{cases} n_x x_c + n_y y_c + n_z z_c = x_w + n_x p_x + n_y p_y + n_z p_z \\ a_x x_c + a_y y_c + a_z z_c = a_x p_x + a_y p_y + a_z p_z \end{cases} \tag{34}$$

Denote the coordinates of a point in the camera frame as (x_{ci}, y_{ci}, z_{ci}). Assume $z_{ci} \neq 0$. Then (34) becomes:

$$a_x x'_{ci} + a_y y'_{ci} + a_z = C_2(n_x x'_{ci} + n_y y'_{ci} + n_z) \tag{35}$$

where $x'_{ci} = x_{ci}/z_{ci}$, $y'_{ci} = y_{ci}/z_{ci}$, and

$$C_2 = \frac{a_x p_x + a_y p_y + a_z p_z}{x_w + n_x p_x + n_y p_y + n_z p_z}.$$

Since vector **n** and **a** are orthogonal and $a_z \neq 0$, it follows that

$$n_x a'_x + n_y a'_y = -n_z \tag{36}$$

where $a'_x = a_x/a_z$ and $a'_y = a_y/a_z$.
Dividing (35) by a_z and eliminating a'_x from (35) and (36) gives:

$$(n_x y'_{ci} - n_y x'_{ci})a'_y - n_x(n_x x'_{ci} + n_y y'_{ci} + n_z)C'_2 = n_z x'_{ci} - n_x \tag{37}$$

where $C'_2 = C_2/a_z$.
As for the points on a line paralleled to the y-axis, their x coordinate, x_w, are the same, and C'_2 should remain constant. Taking any two points on this line gives two equations as (37). Then a'_y and C'_2 can be obtained with these two equations. Substituting a'_y into (36) gives a'_x. Then a_x, a_y and a_z can be determined by normalizing the vector $(a'_x, a'_y, 1)^T$ as (32).
Finally the vector **o** is determined with **o** = **a** x **n**. The rotation matrix is orthogonal since **n** and **a** are unit orthogonal vectors.

4.2.3 Derivation of the coordinates of the translation vector

Taking one point on the line $y = Y_w$ and the other one on the line $y = -Y_w$, the corresponding constants C_1, which are computed with (28), are denoted as C_{11} and C_{12} respectively. Then it follows that

$$\frac{2(o_x p_x + o_y p_y + o_z p_z)}{a_x p_x + a_y p_y + a_z p_z} = C_{11} + C_{12} \tag{38}$$

$$\frac{2Y_w}{a_x p_x + a_y p_y + a_z p_z} = C_{11} - C_{12} \tag{39}$$

Simplifying (38) and (39) gives:

$$\begin{cases} (2o_x - D_{h1}a_x)p_x + (2o_y - D_{h1}a_y)p_y + (2o_z - D_{h1}a_z)p_z = 0 \\ D_{h2}a_x p_x + D_{h2}a_y p_y + D_{h2}a_z p_z = 2Y_w \end{cases} \tag{40}$$

where $D_{h1} = C_{11} + C_{12}$, $D_{h2} = C_{11} - C_{12}$.
Similarly, the line $x = X_w$ and $x = -X_w$ yield that

$$\begin{cases} (2n_x - D_{v1}a_x)p_x + (2n_y - D_{v1}a_y)p_y + (2n_z - D_{v1}a_z)p_z = 0 \\ D_{v2}a_x p_x + D_{v2}a_y p_y + D_{v2}a_z p_z = 2X_w \end{cases} \tag{41}$$

where $D_{v1} = 1/C_{21} + 1/C_{22}$, $D_{v2} = 1/C_{21} - 1/C_{22}$. C_{21} and C_{22} can be computed with (35).
Then the translation vector $\mathbf{p} = (p_x, p_y, p_z)$ can be determined by solving (40) and (41).
Xu et al. gave an improved result of the translation vector **p**, where the area of the rectangle was employed to refine the estimation (Xu et al., 2006b).

4.3 Experiments and results

An experiment was conducted to compare the visual measurement method, which considering the shape constraints, with the traditional stereovision method. A colored rectangle mark was place in front of the humanoid robot. The mark had a dimension of 100mm × 100mm. The parameters of the camera are described in section 3.3.
The edges of the rectangle were detected with Hough transformation after distortion corrections. The intersections between the edges of the rectangle and the x-axis and y-axis of the reference frame were taken as the feature points for stereovision method. The position and orientation of the rectangle relative to the camera reference frame are computed with the Cartesian coordinates of the feature points.
Three measurements were taken under the same condition. Table 3 shows the results. The first column is the results of the traditional stereovision method, while the 2nd column

shows the results of the algorithm presented in section 4.2. It can be found out that the results of the stereovision method were unstable, while the results of the method with the shape constraints were very stable.

Index	Results with stereovision	Results with the proposed method
1	$\begin{bmatrix} 0.4480 & 0.8524 & 0.2550 & 83.6 \\ -0.8259 & 0.2957 & 0.4850 & 63.9 \\ 0.3423 & -0.4313 & 0.8365 & 921.2 \end{bmatrix}$	$\begin{bmatrix} 0.4501 & 0.8397 & 0.3037 & 91.1 \\ -0.8458 & 0.2917 & 0.4467 & 76.4 \\ 0.2865 & -0.4579 & 0.8415 & 959.6 \end{bmatrix}$
2	$\begin{bmatrix} 0.4420 & 0.8113 & 0.3428 & 83.1 \\ -0.8297 & 0.2642 & 0.5071 & 64.3 \\ 0.3409 & -0.5216 & 0.7899 & 918.4 \end{bmatrix}$	$\begin{bmatrix} 0.4501 & 0.8397 & 0.3037 & 91.1 \\ -0.8458 & 0.2917 & 0.4467 & 76.4 \\ 0.2865 & -0.4579 & 0.8415 & 959.6 \end{bmatrix}$
3	$\begin{bmatrix} 0.4480 & 0.8274 & 0.3053 & 83.4 \\ -0.8259 & 0.2811 & 0.5010 & 63.9 \\ 0.3423 & -0.4861 & 0.8093 & 923.3 \end{bmatrix}$	$\begin{bmatrix} 0.4501 & 0.8397 & 0.3037 & 91.1 \\ -0.8458 & 0.2917 & 0.4467 & 76.4 \\ 0.2865 & -0.4579 & 0.8415 & 959.6 \end{bmatrix}$

Table 3. Measuring results for the position and orientation of an object

More experiments were demonstrated by Xu et al. (Xu et al., 2006b). The results indicate that visual measurement with the shape constraints can give a more robust estimation especially when presented with noises in the feature extraction.

5. Hand eye coordination of humanoids robot for grasping

5.1 Architecture for vision guided approach-to-grasp movements

Differ from industrial manipulators, humanoid robots are mobile platforms and the object for grasping can be placed anywhere in the environment. The robot needs to search and approach the object, and then perform grasping with its hand. In this process, both the manipulator and the robot itself need to be controlled. Obviously the required precision in the approaching process is different from that in the grasping process. The requirement for the control method should also be different. In addition, the noises and errors on the system, including the calibration of the vision system, the calibration of the robot, and the visual measurement, will play an important role in the accuracy of visual control (Gans et al., 2002). The control scheme should be robust to these noises and errors.

The approach-to-grasp task can be divided into five stages: searching, approaching, coarse alignment of the body and hand, precise alignment of the hand, and grasping. At each stage, the requirements for visual control are summarized as follows:

1. Searching: wandering in the workspace to search for the concerned target.
2. Approaching: approaching the target from far distance, only controlling the movement of the robot body.
3. Coarse alignment: aligning the body of the robot with the target to ensure the hand of the robot can reach and manipulate the target without any mechanical constrains; also aligning the hand with the target. Both the body and the hand need to be controlled.
4. Precise alignment: aligning the hand with the target to achieve a desired pose relative to the target at a high accuracy. Only the hand of the robot has to be controlled.
5. Grasping: grasping the target based on the force sensor. The control of the hand is needed.

With the change of the stages, the controlled plant and the control method also change. Figure 6 is the architecture of the control system for visual guided grasping task.

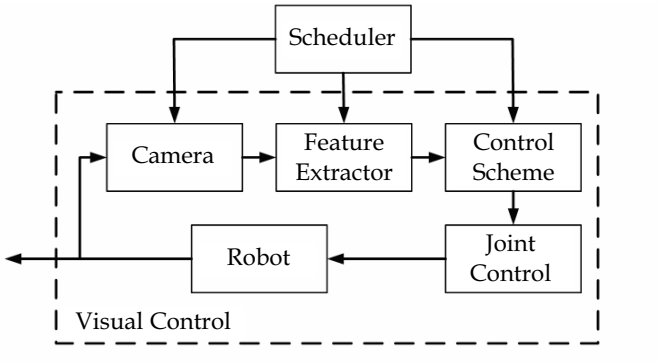


Figure 6. Architecture of the vision control system

In Fig. 6, the visual control block is a basic visual feedback loop. Depending on the used camera-robot configuration and control law, it can be configured as eye-in-hand or eye-to-hand system, with position-based or image-based visual control method. The scheduler module monitors the undergoing of the task. It determines which stage should be carried into execution and invokes correspond image processing and visual control module.

5.2 Valve operation by the humanoid robot

An approach-to-grasp task was designed for the humanoid robot, which is shown in Fig. 1. The robot has a head, a body with two arms and a wheeled mobile base. A stereo rig is mounted on the head as the eyes. Two six DOFs manipulators serve as arms. Each arm has a gripper as the hand. The wrist of the hand is equipped with a mini camera and force sensors.

A valve is placed in the workspace of the humanoid robot. A red rectangle mark is attached on the valve for the robot to identify the valve and estimate its pose. Two green marks are attached on the handles of the valve. The robot will search the valve with its stereo rig on the head. Once the robot finds the valve, it moves towards it and operates it with its hands, as shown in Fig. 7. Operations include turning on and turning off the valve.

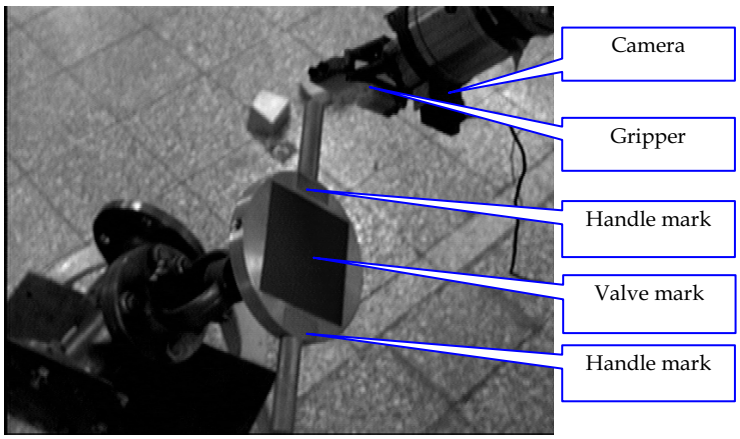


Figure 7. Valve operation with the robot hand

No	Stage	Controlled variables	DOFs	Image features	Camera-robot configuration	Visual control law
1	Searching	${}^{b}\mathbf{T}_{e}$	2 of head 2 of body	Color	EIH	Look-then move
2	Approaching	${}^{b}\mathbf{T}_{e}$	2 of body	Centroid	EIH	Position-based
3	Coarse alignment	${}^{b}\mathbf{T}_{e}$, ${}^{g}\mathbf{T}_{e}$	3 of body 6 of arm	Vertex	EIH, ETH	Position-based
4	Precise alignment	${}^{g}\mathbf{T}_{e}$	4 of arm	Centroid, Area	EIH	Image-based
5	Grasping	${}^{g}\mathbf{T}_{e}$	6 of arm 2 of hand			

Table 4. Summarization of visual control strategy

Table 4 demonstrates the algorithms for each stage of the approach-to-grasp movement. Algorithms for each stage are simple and easy to carry. The methods for calibration and pose estimation discussed in section 3 and section 4 are employed. The complicated approach-to-grasp task such as valve operation can be accomplished with the integration of these methods.

The advantages of both eye-to-hand and eye-in-hand systems are fully exploited in this visual control strategy. The blocking problem for the eye-to-hand system is effectively avoided since cameras on the head are active. The problem of losing targets in the field of view for an eye-in-hand system is resolved, as the hands are only adjusted in a small range.

5.3 Approach and alignment

5.3.1 Approaching

Assume the stereo rig on the head, as well as the head-eye geometry, have been accurately calibrated. After the robot finds the valve based on color with the stereo rig, the center point of the red mark is taken as the image feature. The Cartesian coordinates of the valve can be obtained with the stereovision algorithm using the pixel coordinates of the image feature. A position-based visual control scheme is adopted to control the robot to approach the valve. This is an EIH system if the body of the robot is regarded a "manipulator".

Let the notation ${}^{c}\mathbf{p}_{e}$ represent the position of the mark/valve relative to the camera, ${}^{b}\mathbf{p}_{e}{}^{*}$ represent the desired position of the valve relative to the robot body, which is about 2m away from the robot. An error function (Hutchinson et al., 1996) can be defined as

$$ {}^{b}\mathbf{e} = ({}^{b}e_{x}, {}^{b}e_{y}, {}^{b}e_{z})^{T} = ({}^{b}\mathbf{P}_{e}{}^{*} - {}^{b}\mathbf{P}_{e}) = ({}^{b}\mathbf{P}_{e}{}^{*} - {}^{b}\mathbf{T}_{h} \cdot {}^{h}\mathbf{T}_{c} \cdot {}^{c}\mathbf{p}_{e}) \tag{42} $$

where ${}^{b}\mathbf{T}_{h}$, ${}^{h}\mathbf{T}_{c}$ can be obtained from robot kinematics and camera calibration.

The robot can only move on the ground, which is the xoy plane of the robot base frame as shown in Fig 2. So the z coordinate of ${}^{b}\mathbf{e}$, can be removed from the error. A proportion control law can be designed to eliminate the error, $({}^{b}e_{x}, {}^{b}e_{y})^{T}$, and make the robot approach the valve. In practice, problems such as route plan and obstacle avoidance should also be considered.

5.3.2 Pose estimation and coarse alignment

When the distance between the robot and the valve is near enough, the approaching stage ends, and the coarse alignment starts. Two types of alignment is required. One is the

alignment of robot body with respect to the valve, and the other one is the alignment of the hand with the valve. For the former, the camera-robot configuration is an eye-in-hand system, which is same to the case in approaching. As for the latter, the cameras on the head and the arm form an eye-to-hand system.

The pose of the valve can be estimated with the red rectangle mark. Four vertexes of the mark are taken as the image features. Then the visual measurement method with shape constraints, described in section 4, is employed to measure the pose of the mark, i.e. the pose of the valve, which is denoted by ${}^{c}\mathbf{T}_{e}$.

First, the alignment of the robot body is carried out. The task function can be defined as:

$$ {}^{b}\mathbf{E} = {}^{b}\mathbf{T}_{b^*} = {}^{b}\mathbf{T}_{h} \cdot {}^{h}\mathbf{T}_{c} \cdot {}^{c}\mathbf{T}_{e} \cdot {}^{e}\mathbf{T}_{b^*} \tag{43} $$

where ${}^{c}\mathbf{T}_{e}$ is the estimated pose, ${}^{e}\mathbf{T}_{b^*}$ is the desired pose of the body relative to the valve.

${}^{b}\mathbf{E}$ is a homogeneous transformation matrix. The translation vector $({}^{b}e_x, {}^{b}e_y, {}^{b}e_z)^T$ and the Euler angle $({}^{b}\theta_x, {}^{b}\theta_y, {}^{b}\theta_z)^T$ of ${}^{b}\mathbf{E}$ is taken as the error. Similar to the case of approaching, only $({}^{b}e_x, {}^{b}e_y, {}^{b}\theta_z)^T$ needs to be controlled. Then a position-based visual servoing law is adopted to align the robot body with the valve. The determination of ${}^{e}\mathbf{T}_{b^*}$ should ensure the robot just stand in front of the valve and the hand can reach the valve. The precision of the visual control for the alignment of the body does not need to be high. On the contrary, the movement of the robot can stop once the arm could reach and manipulate the valve.

The alignment of the arm with the valve is a typical position-based visual servoing process. Let:

$$ {}^{g}\mathbf{E} = {}^{g}\mathbf{T}_{g^*} = {}^{g}\mathbf{T}_{b} \cdot {}^{b}\mathbf{T}_{h} \cdot {}^{h}\mathbf{T}_{c} \cdot {}^{c}\mathbf{T}_{e} \cdot {}^{e}\mathbf{T}_{g^*} \tag{44} $$

where ${}^{e}\mathbf{T}_{g^*}$ is the desired pose of the end-effector relative to the valve, ${}^{c}\mathbf{T}_{e}$ is estimated with the stereo rig, ${}^{h}\mathbf{T}_{c}$ represents the head-eye geometry, ${}^{b}\mathbf{T}_{h}$ is the transformation from the head to the robot base, and ${}^{g}\mathbf{T}_{b}$ is the relative pose between the hand and the robot base frame. ${}^{b}\mathbf{T}_{h}$ and ${}^{g}\mathbf{T}_{b}$ can be obtained with the robot kinematics.

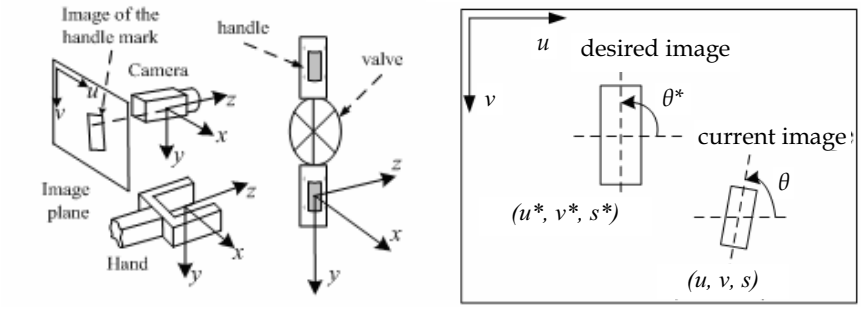
Since ${}^{g}\mathbf{E}$ is relevant to ${}^{c}\mathbf{T}_{e}$, ${}^{g}\mathbf{T}_{b}$, ${}^{b}\mathbf{T}_{h}$ and ${}^{h}\mathbf{T}_{c}$, it would be sensitive to noises and errors in robot kinematics, vision system calibration and pose estimation. It is hard to achieve a high accurate alignment, so this stage is called coarse alignment.

The image obtained from the camera mounted on the wrist is checked while the hand approaching the valve. The coarse alignment will stop if the pixel area of the green mark is large enough or the goal pose ${}^{e}\mathbf{T}_{g^*}$ has been reached.

5.4 Image based visual control for accurate manipulation

In this stage, the arm and the camera on the hand form an eye-in-hand system. an image-based visual servoing method is employed to accurately align the hand with the valve. Figure 8 (a) demonstrates the coordinate frames between the hand and the valve. Assume the valve is in front of the hand, and the orientations of these two frames are the same. The DOFs needed to be controlled for the hand are the translation and the rotation around the z-axis in the hand frame, denoted as $\mathbf{r} = (x, y, z, \theta_z)^T$.

With the green mark on the handle, the pixel coordinates of the center of the mark, the pixel area of the mark, and the angle between the principle axis of the mark and the horizontal direction on the image are taken as the image features, which is denoted as $\mathbf{f} = (u, v, s, \theta)^T$. Figure 8 (b) is the sketch of the image features. The desired image features, $\mathbf{f^*} = (u^*, v^*, s^*, \theta^*)^T$, are obtained off-line by means of teach-by-showing.



(a) (b)

Figure 8. Image-based visual servoing, (a) Frame assignment, (b) Image feature

Assume the camera has been mounted on the wrist in such a manner that the orientation of the camera frame is almost the same as that of the hand frame, as shown in Fig. 8 (a). Then the image will change horizontally when the hand moves along the x-axis of the hand frame, and the image will change vertically when the hand moves along the y-axis. Then the image feature vector, **f**, can be taken as an estimation of **r**. The error is defined as follows:

$$\mathbf{e}(\mathbf{f}) = (u^* - u, v^* - v, (s / s^*) - 1, \theta^* - \theta)^T \tag{45}$$

Then a decoupled proportional control law can be designed:

$$(\Delta x, \Delta y, \Delta z, \Delta\theta_z)^T = -(K_u(u^* - u), K_v(v^* - v), K_z((s/s^*) - 1), K_\theta(\theta^* - \theta))^T \tag{46}$$

where K_u, K_v, K_z and K_θ are the proportional parameters of the controller.

5.5 Experiments and results

Valve operation experiments were conducted with the humanoid robot shown in Fig. 1. The size of the red rectangle mark was 100mm x 100mm. The two handles of the valve were attached with green marks. The head with two MINTRON 8055MK cameras is shown in Fig. 2. Two mini cameras were mounted on the wrists as shown in Fig. 7. The head-eye geometry was well calibrated with the method in section 3.

The size of the image obtained from the stereo rig is 768 x 576 pixels. The parameters of the stereo rig are shown in Table 5 and the displacement between the cameras is given in (47).

Item	Left Camera	Right Camera
K_u	4.2180e-007	3.4849e-007
K_v	3.6959e-007	3.6927e-007
K_x	1.0823e+003	1.2522e+003
K_y	1.0678e+003	1.2423e+003
u	324.6	335.0
v	248.3	292.0

Table 5. Parameters of the stereo rig

$$ {}^{R}T_{L} \begin{bmatrix} 0.9998 & -0.0198 & -0.0063 & -93.4482 \\ 0.0195 & 0.9990 & -0.0403 & 1.4111 \\ 0.0071 & 0.0402 & 0.9991 & 125.7946 \end{bmatrix} \tag{47}$$

Firstly, the robot searched for the valve in the laboratory. When the valve was found, the approaching stage described in section 5.3 started. When the valve was within two meters from the robot, the coarse alignment began. The position of the mobile base of the robot was adjusted according to the pose of the valve until the hand could reach the valve. When the robot stopped moving, the position and orientation of the valve was measured by the stereo rig on the robot head. Table 6 shows the pose of the valve relative to the base frame of the robot, which is attached at the chest. In the process of approaching the valve, two arms were positioned so that they did not block the stereo rig to observe the valve. A pair of images obtained from the stereo rig at the end of the coarse alignment is shown in Fig. 9. It is shown that the hand was at the place near to the handle with an appropriate pose.

The hands of two arms would move to the two handles of the valve respectively. At the same time the cameras on the head were inactive, while the camera at each hand was in operation to observe the green mark on the valve handle. Then accurate alignment of the hand was started. The hands of robot were guided to the handle stably and accurately. Finally, a hybrid control method using force and position was employed to rotate the valve after the hands reached the handle and grasped it successfully.

In a series of experiments, the humanoid robot was able to autonomously find, approach and operate the valve successfully. The advantages of combining both eye-to-hand and eye-in-hand systems are clearly demonstrated.

n	o	a	p(mm)
-0.3723	-0.6062	-0.6795	-905.8
0.0279	0.7225	-0.6862	-36.7
0.9277	-0.3326	-0.2522	124.5

Table 6. Position and orientation of the valve

(a) (b)

Figure 9. Images obtained with the stereo rig on the robot head, (a) Left, (b) Right

6. Conclusion

Issues concerning with the approach-to-grasp movement of the humanoid robot are investigated in this chapter, including the calibration of the vision system, the visual measurement of rectangle objects and the visual control strategy for grasping.

A motion based method for head-eye calibration is proposed. The head-eye geometry is determined with a linear equation and then refined by a nonlinear least squares optimization to ensure the orthogonality of the rotation matrix.
A visual measurement algorithm is provided for the pose estimation of a rectangle object. Both the accuracy of the measurement and the robustness to noises in feature extraction are improved with the shape constraints of the rectangle object
A visual control strategy is presented in this chapter, which integrates different visual control method to fulfil the complicated approach-to-grasp task. The whole process of the grasping task is divided into several stages. The precision requirement of each stage is matched with an appropriate visual control method. Eye-to-hand and eye-in-hand architectures are integrated, at the same time, position-based and image-based visual control methods are combined to achieve the grasping.
A valve operating experiment with a humanoid robot was conducted to verify these methods. The results show that the robot can approach and grasp the handle of the valve automatically with the guidance of the vision system.
Vision is very important for humanoid robots. The approach-to-grasp movement is a basic but complex task for humanoid robots. With the guidance of the visual information, the grasping task can be accomplished. The errors on calibration of the vision system and the robot system will affect the accuracy of the visual measurement and the visual control. Improving the robustness of algorithms to these errors and noises should be the efforts in the future work. In addition, methods for stably identifying and tracking objects in an unstructured environment also need to be studied.

7. Acknowledgement

The authors would like to acknowledge the National High Technology Research and Development Program of China (grant No. 2006AA04Z213), the National Key Fundamental Research and Development Project of China (grant No. 2002CB312200) and the National Natural Science Foundation of China (grant No. 60672039) for the support to this work. The authors would also like to acknowledge the United Laboratory of Institute of Automation, Chinese Academy of Sciences (CASIA) and University of Science and Technology of China (USTC) for the support to this work.

8. References

Chaumette, F. & Hutchinson, S. (2006). Visual servo control Part I: Basic approaches. *IEEE Robotics and Automation Magazine*, Vol. 13, No. 4, pp. 82-90, ISSN: 1070-9932

Clarke, T. A. & Fryer, J. G. (1998). The development of camera calibration methods and models. *Photogrammetric Record*, Vol. 16, No. 91, pp. 51-66, ISSN: 0031-868X

Corke, P. I. & Hutchinson, S. A. (2001). A new partitioned approach to image-based visual servo control. *IEEE Transactions on Robotics and Automation*, Vol. 17, No. 4, pp. 507-515, ISSN: 1042-296X

Dodds, Z.; Jagersand, M; Hager, G. D & Toyama, K. (1999). A hierarchical vision architecture for robotic manipulation tasks, *Proceedings of First International Conference on Computer Vision System*, pp. 312-330, ISSN: 0302-9743, Las Palmas, Gran Canaria, Spain, Jan. 1999, Springer, Berlin, German

Faugeras, O. D. & Toscani, G. (1986). The calibration problem for stereo. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 15-20, ISBN: 0-8186-0721-1, Miami Beach, USA, 1986, IEEE Press, New York, USA

Flandin, G.; Chaumette, F. & Marchand, E. (2000). Eye-in-hand/Eye-to-hand cooperation for visual servoing. *Proceedings of 2000 IEEE International Conference on Robotics and Automation,* pp. 2741-2746, ISBN: 0-7803-5886-4, San Francisco, California, USA, Apr. 2000, IEEE Press, New York, USA

Gans, N. R.; Corke, P. I & Hutchinson, S. A. (2002). Performance tests of partitioned approaches to visual servo control, *Proceedings of 2002 IEEE International Conference on Robotics and Automation*, pp. 1616-1623, ISBN: 0-7803-7272-7, Washington, DC, USA, May 2002, IEEE Press, New York, USA

Hager, G. D.; Chang, W. C. & Morse, A. S. (1995). Robot hand-eye coordination based on stereo vision. *IEEE Control Systems Magazine*, Vol. 15, No. 1, pp. 30-39, ISSN: 0272-1708

Han, M.; Lee, S.; Park, S. K. & Kim, M. (2002). A new landmark-based visual servoing with stereo camera for door opening, *Proceedings of International Conference on Control, Automation and Systems*, pp. 1892-1896, Muju Resort, Jeonbuk, Koreaa, Oct. 2002

Hartley, R. & Zisserman, A. (2004). *Multiple view geometry in computer vision (Second Edition)*, Cambridge University Press, ISBN: 0521540518, London, UK

Hashimoto, K.; Kimoto, T; Ebine, T. & Kimura, H. (1991). Manipulator control with image-based visual servo, *Proceedings of 1991 IEEE International Conference on Robotics and Automation*, pp. 2267-2272, ISBN: CH2969-4, Sacramento, California, USA, Apr. 1991, IEEE Press, New York, USA

Hauck, A.; Sorg, M.; Farber, G. & Schenk, T. (1999). What can be learned from human reach-to-grasp movements for the design of robotic hand-eye system?. *Proceedings of 1999 IEEE International Conference on Robotics and Automation*, pp. 2521-2526, ISBN: 0-7803-5180-0-5, Detroit, Michigan, USA, May 1999, IEEE Press, New York, USA

Hauck, A.; Passig, G. Schenk, T. Sorg, M. & Farber, G. (2000). On the performance of a biologically motivated visual control strategy for robotic hand-eye coordination, *Proceedings of 2000 IEEE/RSJ International Conference on Intelligent Robotics and Systems*, pp. 1626-1632, ISBN: 0-7803-6348-5, Takamatsu, Japan, Oct. 2000, IEEE Press, New York, USA

Heikkila, J. & Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. *Proceedings of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1106-1112, ISBN: 1063-6919, San Juan, Puerto Rico, Jun. 1997, IEEE Press, New York, USA

Horaud, R.; Dornaika, F. & Espian, B. (1998). Visually guided object grasping. *IEEE Transactions on Robotics and Automation*, Vol. 14, No. 4, pp. 525-532, ISSN: 1042-296X

Hutchinson, S.; Hager, G. D. & Corke, P. I. (1996). A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, Vol. 12, No. 5, pp. 651-670, ISSN: 1042-296X

Kragic, D.; Miller, A. T. & Allen, P. K. (2001). Real-time tracking meets online grasp planning, *Proceedings of IEEE International Conference on Robotics and Automation*, pp. 2460-2465, ISSN: 1050-4729, Seoul, Korea, May 2001, IEEE Press, New York, USA

Kragic, D. & Christensen, H. I. (2002). Survey on visual servoing for manipulation. *Technical report, Computational Vision and Active Perception Laboratory(CVAP)*, Jan. 2002, CVAP259, ISRN KTH/NA/P--02/01--SE

Kragic, D. & Christensen, H. I. (2003). A framework for visual servoing. *Proceedings of International Conference on Computer Vision Systems*, pp. 345-354, ISSN: 0302-9743, Graz, Austria, Apr. 2003, Springer, Berlin, German

Ma, L.; Chen Y. Q. & Moore, K. L. (2003). Flexible camera calibration using a new analytical radial undistortion formula with application to mobile robot localization. *Proceedings of 20003 IEEE International Symposium on Intelligent Control*, pp. 799-804, ISBN: 0-7803-7891-1, Houston, Texas, USA, Oct. 2003, IEEE Press, New York, USA

Ma, S. D. (1996). A self-calibration technique for active vision system. *IEEE Transactions on Robotics and Automation*, Vol. 12, No. 1, pp. 114-120, ISSN: 1042-296X

Malis, E.; Chaumette, F. & Boudet, S. (1999). 2½D visual servoing. *IEEE Transactions on Robotics and Automation*, Vol, 15, No. 2, pp. 238-250, ISSN: 1042-296X

Murray, R. M.; Li, Z. X. & Sastry, S. S. (1993). *A mathematical introduction to robotic manipulation*, CRC Press, ISBN: 0-8493-7981-4, Boca Raton, Florida, USA

Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Transactions on Robotics and Automation*, Vol. 3, No. 4, pp. 323-344, ISSN: 0882-4967

Tsai, R. Y. & Lenz, R. K. (1989). A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, Vol. 5, No. 3, pp. 345-358, ISSN: 1042-296X

Xu, D.; Li, Y. F. & Tan, M. (2006a). Method for calibrating cameras with large lens distortion. *Optical Engineering*, Vol. 45, No. 4, pp. 043602-1-043602-8 , ISSN: 0091-3286

Xu, D.; Tan, M.; Jiang Z. M. & Hu, H. S. (2006b). Use of colour and shape constraints in vision-based valve operation by robot. *International Journal of Advanced Robotic Systems*, Vol. 3, No. 3, pp. 267-274, ISSN: 1729-8806

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, pp. 1330-1334, ISSN: 0162-8828