

# Use of Colour and Shape Constraints in Vision-based Valve Operation by Robot

**De Xu<sup>1</sup>; Min Tan<sup>1</sup>; Zemin Jiang<sup>1</sup> & Huosheng Hu<sup>2</sup>**

<sup>1</sup>Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P.R. China, xude@compsys.ia.ac.cn

<sup>2</sup>Department of Computer Science, University of Essex, Colchester CO4 3SQ, United Kingdom, hhu@essex.ac.uk

**Abstract:** *This paper proposes a new strategy for a humanoid robot to approach and operate a valve based on colour and shape constraints. It consists of four stages, namely rough base approaching, fine base approaching, rough hand approaching and fine hand approaching and grasping. The robot estimates the object's position using its stereo-vision at the first stage. A new visual positioning method is used to determine the valve's position and pose in the robot's frame in the second stage. When its hands are near the valve, a visual servoing method is employed to catch the handle of the valve via cameras in end-effectors. The advantages of both eye-in-head and eye-to-hand systems are exploited. Experimental results are presented to verify the effectiveness of the proposed method.*

**Keywords:** *Visual positioning, visual control, hand-eye system, autonomous manipulation, humanoid robot.*

## 1. Introduction

Hand-eye systems is widely used in robotics applications, which include two types: one is an eye-in-hand system (EIHS) that has cameras installed on and moved with hands, and the other is an eye-to-hand system (ETHS) that has cameras that do not move with hands (Flandin 2000).

EIHS is very popular in the area of industrial robotics. When a manipulator approaches a target, the distance between the camera and the target is reduced, and the measurement error of the camera is decreased. Visual control methods in an EIHS are divided into three types, namely image-based, position-based or hybrid (a combination of both). The image-based visual control method can effectively eliminate camera calibration error because of the closed loop established in the image space. On the other hand, the absolute measurement error in position-based visual control is dramatically reduced while a manipulator is close to a target. The same situation happens under the hybrid visual control method (Hager 1996, Chaumette 2000, Corke 2000, Zhu 2000, Wells 2001). However, EIHS has a vital drawback, i.e. the object cannot be guaranteed in the view field of the cameras at all time, especially during the pose adjustment of the hand at a long range (Hager 1996).

In contrast, ETHS can be effectively used in humanoid robots and mobile manipulators that operate in a large work space. When the robot is far from a target, it travels toward it and stops at a close range. Then, according to visual measurements, the manipulator approaches the

target and manipulates it. To ensure that the end-effector can reach the target accurately, some researchers have designed special marks which are installed on the end-effector and the target (Han 2002, Cardenas 2003). The approaching task is realized through closed loop control of end-effectors. However, because the target or markers may be partially blocked during approaching or manipulation, image-based or hybrid visual control methods may not be able to bring the manipulator to the target accurately.

As we know, the position of an object in 3D space can be calculated from two image points using stereo cameras and according to the projecting view lines. The lack of constraints, errors in calibration and errors in image coordinates of matching points result in large errors during object positioning and pose estimation. By using shape constraints of an object and its multiple imaging points, positioning accuracy, especially pose estimation accuracy, can be increased and the influence of the last factor can be partly eliminated (Bartoli 2001). By combining ETHS and EIHS, a humanoid robot could use its hands to reach and manipulate an object accurately.

In this paper, the advantages of both eye-to-hand and eye-in-hand systems are fully exploited in the development of a new positioning method. 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, and end-effectors only adjust their position in a small range. The rest of this paper is organized as follows. Section 2 introduces our humanoid robot and the

four-stage process for finding and manipulating a valve. The camera models are described in Section 3. Section 4 proposes a new visual positioning method based on rectangle constraints, which accurately provides the position and pose of the valve. System calibration is conducted in Section 5 to verify the accuracy of the proposed positioning method. Section 6 presents the application experiment that is designed for the humanoid robot to approach and operate the valve autonomously, and results show the effectiveness of the proposed method. Finally, Section 7 provides a brief conclusion.

## 2. The robot and its control strategy

As shown in Fig. 1, our humanoid robot consists of a head, a body with two arms and a wheeled mobile base. The robot body has three degrees of freedom (DOFs), i.e. twist, pitch and yaw. The two arms/manipulators have six DOFs and are fixed, one on each side of the body. Each has an end-effector as its hand, and its wrist is equipped with a mini camera and force sensors. Note that we treat the end-effect, gripper and hand as the same in this paper from now on without further explanation.

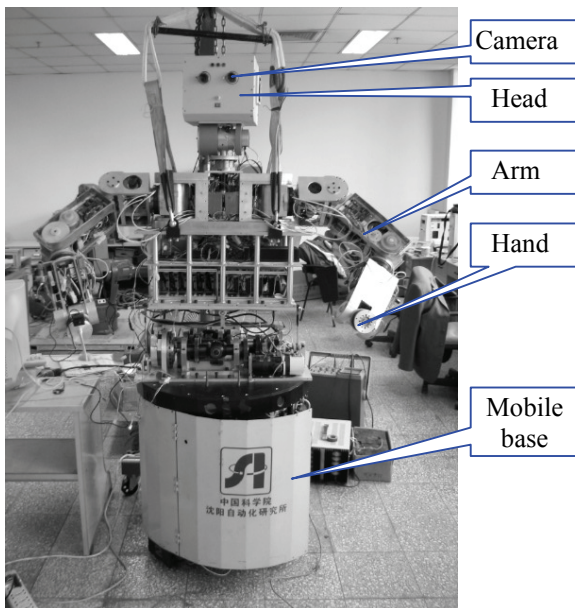


Fig. 1. The humanoid robot

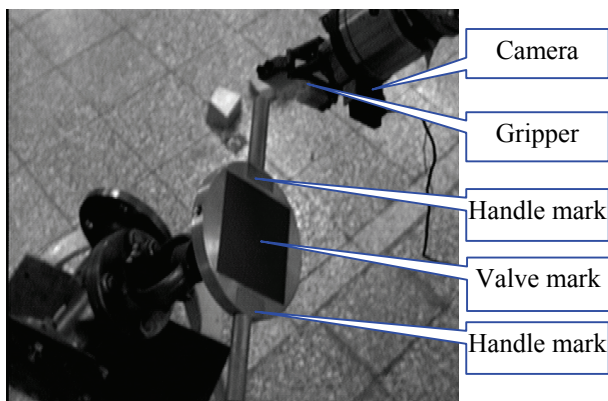


Fig. 2. Valve with a rectangle mark

The robot head has two cameras as eyes and a PC104 computer to process images used to position the valve. Once the robot finds the valve, it moves towards it and operates it using its hands, as shown in Fig. 2. Operations include turning on or turning off the valve. These operations can be remotely controlled by an operator using audio commands sent via radio.

The process of finding and operating the valve consists of four stages as follows:

- Stage 1 – The robot first uses its stereo vision to estimate the rough position of the valve relative to its own position in order to approach the valve. At this stage, the centre of the image area for the red colour marker is selected as the feature point, and the pose of the valve is not important. When the distance between the valve and the robot is less than two meters, the first stage of the positioning method is ended and the second stage begins. A new strategy is developed for measuring the position and pose of the valve, in the robot frame, based on the shape constraint of the marker.
- Stage 2 – According to the position and pose of the valve in the robot frame, the robot moves near to it at a range that is reachable by its arm. The position and pose of the valve, calculated at the end of the 2nd stage, is used for the movement control of its arm in the 3rd stage of the positioning method. The given pose of the end-effector of the robot arm is calculated, and is kept for later stages.
- Stage 3 – The position that the hand should reach at this stage is calculated according to the position of the valve (by considering the position of the mark and handles). Based on kinematics and inverse kinematics, the hand is controlled to move to the handle while the camera in the hand measures the green colour image size of the handle marker. It will stop when the marker size is large enough or a given position is reached.
- Stage 4 – An image-based visual servoing method is adopted to guide the end-effector to reach and catch the handle. Finally hybrid control with force and position is employed to rotate the valve using two hands.

With regard to control, many methods are employed in the process described above. The control methods in the 1st and 2nd stages employ position-based visual servoing. Control in the 3rd stage employs models and control in the 4th stage involves image-based visual servoing. The position-based visual servoing methods in the 1st and 2nd stage and the model based control method in the 3rd stage are traditional. They are omitted here because of length limitation. The pose of the valve, given at the end of the 2nd stage, is an important parameter because it ensures that the end-effector can catch the handle with correct orientation. The visual positioning method in the 2nd stage will be described in the next section.

### 3. Camera model

To enlarge the field of view, 8mm focus lenses are selected for the cameras in the robot head. However, this kind of lenses has the distortion problem, which needs to be corrected. In this research, the process of distortion correction is carried out by simply changing a non-linear image to a linear one. In other words, the imaging curve of a line should be corrected to a linear line. To simplify the process, the non-linear model shown in (1) is used to denote the radial distortion.

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

where  $[u, v]$  are the coordinates of a point in a practical image.  $[u_0, v_0]$  denote the image coordinates of the centre of the optical axis.  $[u', v']$  are the coordinates in the image after distortion correction, and  $[k_u, k_v]$  are the one order correction coefficients of the radial distortion in  $u$  and  $v$  directions.  $r = \sqrt{(u - u_0)^2 + (v - v_0)^2}$  is the radius.

The intrinsic and extrinsic parameter models of the cameras are shown in (2) and (3).

$$\begin{bmatrix} u_d \\ v_d \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c / z_c \\ y_c / z_c \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} x_c / z_c \\ y_c / z_c \\ 1 \end{bmatrix} \quad (2)$$

where  $u_d = u' - u_0$ ,  $v_d = v' - v_0$ , and  $[x_c, y_c, z_c]$  are the coordinates of a point in the camera frame.  $M_1$  is the intrinsic parameter matrix, and  $[k_x, k_y]$  are the magnification coefficients from the imaging plane coordinates to the image coordinates.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_2 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (3)$$

where  $[x_w, y_w, z_w]$  are the coordinates of a point in the object frame, and  $M_2$  is the extrinsic parameter matrix. In  $M_2$ ,  $\bar{n} = [n_x \ n_y \ n_z]^T$  is the direction vector of the  $x$ -axis,  $\bar{o} = [o_x \ o_y \ o_z]^T$  is the direction vector of the  $y$ -axis,  $\bar{a} = [a_x \ a_y \ a_z]^T$  is the direction vector of the  $z$ -axis, and  $\bar{p} = [p_x \ p_y \ p_z]^T$  is the position vector.

### 4. Visual positioning method

A red rectangular colour marker is attached to the valve, as shown in Fig. 2. The measurement of the position and pose for the valve is similar to that for the red marker. A frame is established as a target frame based on the rectangle centre, which takes the rectangle plane as a XOY plane. The line between two handle markers acts as the  $X$  axis, as shown in Fig. 3. The rectangle size is  $2X_w$  in width and  $2Y_w$  in height. The coordinates of the four

vertexes  $P_1$  to  $P_4$  are also known in this frame. Obviously, any point on the plane should satisfy  $z_w=0$ .

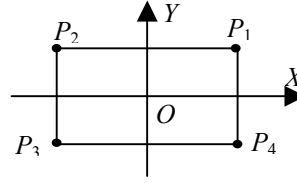


Fig. 3. The objective frame of a rectangle

#### 4.1 The derivation of vector $\bar{n}$

According to the orthogonal constraints of  $M_2$ , we have (4) that is derived from (3) with the condition  $z_w=0$ .

$$\begin{cases} o_x x_c + o_y y_c + o_z z_c = y_w + o_x p_x + o_y p_y + o_z p_z \\ a_x x_c + a_y y_c + a_z z_c = a_x p_x + a_y p_y + a_z p_z \end{cases} \quad (4)$$

Let

$$\begin{cases} A_1 = y_w + o_x p_x + o_y p_y + o_z p_z \\ B_1 = a_x p_x + a_y p_y + a_z p_z \end{cases} \quad (5)$$

Since  $A_1 \neq 0$ ,  $B_1 \neq 0$  and  $z_c \neq 0$ , we have

$$\frac{o_x x'_c + o_y y'_c + o_z}{a_x x'_c + a_y y'_c + a_z} = C_1 \quad (6)$$

where  $C_1 = A_1/B_1$ ,  $x'_c = x_c/z_c$  and  $y'_c = y_c/z_c$ . Note that  $y'_c$  can be obtained from (1) and (2) according to the imaging coordinates  $[u, v]$ .

All points on the line parallel to the  $X$  axis have the same coordinate  $y_w$ , so  $A_1$  and  $B_1$  are constants. Taking two points on the line arbitrarily, such as point  $i$  and point  $j$ , and applying them to (6), we obtain (7) and also its simplified form (8) which results from its simplification using the orthogonal restriction of the rotation matrix  $M_2$ .

$$\frac{o_x x'_{ci} + o_y y'_{ci} + o_z}{a_x x'_{ci} + a_y y'_{ci} + a_z} = \frac{o_x x'_{cj} + o_y y'_{cj} + o_z}{a_x x'_{cj} + a_y y'_{cj} + a_z} \quad (7)$$

$$n_x (y'_{ci} - y'_{cj}) + n_y (x'_{cj} - x'_{ci}) + n_z (x'_{ci} y'_{cj} - x'_{cj} y'_{ci}) = 0 \quad (8)$$

Any two points on the same line parallel to the  $X$  axis should match (8). Therefore, we can obtain two equations for one camera from the two lines parallel to the  $X$  axis, i.e. four such equations for two cameras. If the camera's optical axis is not vertical to the target plane, then  $n_z \neq 0$ , (8) can be divided by  $n_z$  and become (9).

$$n'_x (y'_{ci} - y'_{cj}) + n'_y (x'_{cj} - x'_{ci}) = x'_{cj} y'_{ci} - x'_{ci} y'_{cj} \quad (9)$$

where  $n'_x = n_x/n_z$  and  $n'_y = n_y/n_z$ .  $n_z$  is calculated with  $\|\bar{n}\|=1$  after calculating  $n_x/n_z$  and  $n_y/n_z$ . Then  $n_x$  and  $n_y$  can be obtained easily.

If the camera's optical axis is vertical to the target plane, we have  $n_z=0$  and (8) has two unknown variables. (10) is obtained with  $\|\bar{n}\|=1$  after transposition and the square

of (8), with the results of  $n_x$  and  $n_y$ .  $n_x$  is positive, and the sign of  $n_y$  depends on (8).

$$\begin{cases} n_x^2 = \frac{(x'_{ci} - x'_{ci})^2}{(y'_{ci} - y'_{ci})^2 + (x'_{ci} - x'_{ci})^2} \\ n_y^2 = 1 - n_x^2 \end{cases} \quad (10)$$

#### 4.2. The derivation of vector $\bar{a}$

According to (3) and the orthogonal restriction of the rotation matrix  $M_2$ , we have

$$\begin{cases} n_x x_c + n_y y_c + n_z z_c = A_2 \\ a_x x_c + a_y y_c + a_z z_c = B_1 \end{cases} \quad (11)$$

where  $A_2 = x_w + n_x p_x + n_y p_y + n_z p_z$ .

In a line parallel to the Y axis,  $x_w$  is constant, so  $A_2$  is a constant too. Because  $A_2 \neq 0$ ,  $B_1 \neq 0$  and  $z_c \neq 0$ , (11) becomes

$$a_x x'_{ci} + a_y y'_{ci} + a_z = C_2 (n_x x'_{ci} + n_y y'_{ci} + n_z) \quad (12)$$

where  $C_2 = B_1 / A_2$ .

Considering that  $\bar{n}$  and  $\bar{a}$  are orthogonal, and  $a_z \neq 0$ , we have

$$n_x a'_x + n_y a'_y = -n_z \quad (13)$$

where  $a'_x = a_x / a_z$  and  $a'_y = a_y / a_z$ .

If (12) is divided by  $a_z$ , and then combined with (13),  $a'_x$  is removed. A new equation with variables  $a'_y$  and  $C'_2$  is formed as follows:

$$(n_x y'_{ci} - n_y x'_{ci}) a'_y - n_x (n_x x'_{ci} + n_y y'_{ci} + n_z) C'_2 = n_z x'_{ci} - n_x \quad (14)$$

where  $C'_2 = C_2 / a_z$ .

As for points on the line parallel to the axis Y,  $C'_2$  is a constant. Taking any two points from the line, we calculate both  $a'_y$  and  $C'_2$  by using (14). By applying  $a'_y$  to (13),  $a'_x$  is calculated, followed by the results of  $a_x$ ,  $a_y$  and  $a_z$ , with  $\|\bar{a}\| = 1$ .

To improve accuracy, points on two lines parallel to the Y axis are used to calculate the results of  $a_x$ ,  $a_y$  and  $a_z$ . It should be noticed that each line parallel to the Y axis has a different constant  $C'_2$  in (14).  $\bar{o}$  is determined from  $\bar{o} = \bar{a} \times \bar{n}$  after obtaining the vectors  $\bar{n}$  and  $\bar{a}$ . The rotation matrix can be ensured as an orthogonal matrix since  $\bar{n}$  and  $\bar{a}$  are unit orthogonal vectors.

#### 4.3. The derivation of vector $\bar{p}$

##### 1) Rough Positioning:

Two points are taken from two lines, one on line  $y = Y_w$  and the other on line  $y = -Y_w$ . The corresponding constants  $C_1$  for the lines are calculated from (6), denoted as  $C_{11}$  and  $C_{12}$  respectively.  $C_{11}$  corresponds to  $y = Y_w$ . To enhance the precision, more points on each line are needed.  $C_{11}$  and  $C_{12}$  are calculated from the average of  $C_1$  for each line respectively. From  $C_{11}$  and  $C_{12}$ , two

equations (15) and (16) with  $p_x$ ,  $p_y$  and  $p_z$  are obtained. Equation (17) is derived from them.

$$\frac{2(o_x p_x + o_y p_y + o_z p_z)}{a_x p_x + a_y p_y + a_z p_z} = C_{11} + C_{12} \quad (15)$$

$$\frac{2Y_w}{a_x p_x + a_y p_y + a_z p_z} = C_{11} - C_{12} \quad (16)$$

$$\begin{cases} (2o_x - D_{h1} a_x) p_x + (2o_y - D_{h1} a_y) p_y \\ + (2o_z - D_{h1} a_z) p_z = 0 \\ D_{h2} a_x p_x + D_{h2} a_y p_y + D_{h2} a_z p_z = 2Y_w \end{cases} \quad (17)$$

where  $D_{h1} = C_{21} + C_{22}$ ,  $D_{h2} = C_{21} - C_{22}$ .

Similarly, two equations are formed from line  $x = X_w$  and  $x = -X_w$  as in (18). Combining (17) and (18),  $p_x$ ,  $p_y$  and  $p_z$  are calculated.

$$\begin{cases} (2n_x - D_{v1} a_x) p_x + (2n_y - D_{v1} a_y) p_y \\ + (2n_z - D_{v1} a_z) p_z = 0 \\ D_{v2} a_x p_x + D_{v2} a_y p_y + D_{v2} a_z p_z = 2X_w \end{cases} \quad (18)$$

where  $D_{v1} = 1/C_{21} + 1/C_{22}$ ,  $D_{v2} = 1/C_{21} - 1/C_{22}$ .  $C_{21}$  and  $C_{22}$  correspond to lines  $x = X_w$  and  $x = -X_w$  respectively.

##### 2) Fine Positioning:

In the camera frame,  $\bar{P}$  is a position vector of the original point in the target frame. Obviously, using  $\bar{P}$  and (2), the image coordinates of the target original point,  $[u_b, v_b]$ , are obtained easily.

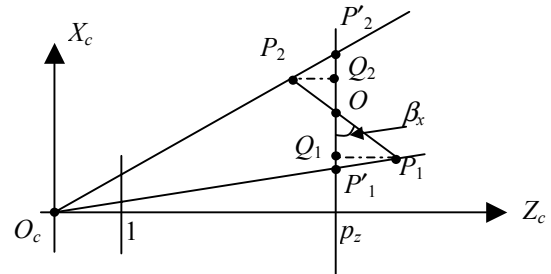


Fig. 4. Space position and imaging

Fig. 4 shows the relation between the space point and its imaging point. According to the camera's pinhole model, the target point  $P_1$  in the space and the point  $P'_1$  on the plane  $Z_c = p_z$  share the same imaging coordinates. Using the imaging coordinates  $P'_1$  and angle  $\beta_x$ , the  $x$ -coordinate of the point  $P_1$  in the target frame is obtained as follows:

$$P_{1x} = OP'_1 = \frac{OP'_1}{\cos \beta_x + (P'_{1x} / p_z) \sin \beta_x} \quad (19)$$

where  $\beta_x$  is the angle between projections of the Z axis in the target frame and the  $Z_c$  axis in the image frame on the plane  $X_c O_c Z_c$ , as shown in (20).  $P'_{1x} / p_z = x_{c1} / z_{c1}$  is obtained from (2) using the imaging coordinates.  $OP'_1$  is the offset on the  $X_c$  axis between the points  $P'_1$  and  $O$  on the plane  $Z_c = p_z$ , as in (21).

$$\beta_x = a \tan 2(a_x, a_z) \quad (20)$$

$$OP'_1 = \frac{u_1 - u_b}{k_x} p_z \quad (21)$$

Applying (20) and (21) to (19),  $P_{1x}$ , the coordinate for  $P_1$  on the axis X in the target frame is calculated.

$$P_{1x} = \frac{(u_1 - u_b)\sqrt{a_x^2 + a_z^2}}{a_z k_x + a_x(u_1 - u_0)} p_z = m_{1x} p_z \quad (22)$$

Similarly, the coordinate for  $P_1$  on the Y axis in the target frame,  $P_{1y}$ , is also obtained.

$$P_{1y} = \frac{(v_1 - v_b)\sqrt{a_y^2 + a_z^2}}{a_z k_y + a_y(v_1 - v_0)} p_z = m_{1y} p_z \quad (23)$$

where  $m_{1x}$  and  $m_{1y}$  in (22) and (23) are the coordinates of  $P_1$  in the virtual target frame, which is in the normalized focus imaging plane, i.e.  $p_z=1$ .

In the target plane, coordinates offset on the axis Y of both top brim and bottom brim of the rectangle are integrated along the axis X, obtaining the area  $S$  of the target rectangle.

$$S = \sum_{i=1}^M (P_{2y}^i - P_{1y}^i)(P_{1x}^{i+1} - P_{1x}^i) \quad (24)$$

$$= [\sum_{i=1}^M (m_{2y}^i - m_{1y}^i)(m_{1x}^{i+1} - m_{1x}^i)] p_z^2 = S_1 p_z^2$$

where  $S_1$  is the target area on the normalized focus imaging plane.  $M$  is the sample numbers along the X axis in the target frame.  $m_{1y}^i, m_{2y}^i$  are coordinates on the Y axis of the  $i$ -th edge point of the rectangle, which have the same coordinates on the X axis.

$$p_z = \sqrt{S/S_1} = 2\sqrt{X_w Y_w / S_1} \quad (25)$$

$p_x$  and  $p_y$  are recalculated by applying  $p_z$  to (17) and (18), to obtain more precise target position vector  $\bar{p}$ .

## 5. System calibration

### 5.1. Camera Calibration

The two cameras on the robot head were well calibrated using the method described in (Zhang 2000, Heikkela 2000). Their intrinsic parameters are shown in Table 1. The extrinsic parameters of the left camera relative to the end of the industrial robot are given in (26).

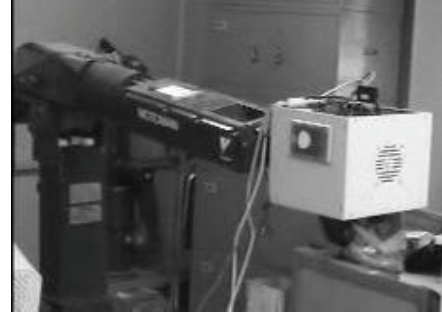
$$T_m = \begin{bmatrix} -0.0452 & -0.6712 & -0.7399 & 47.9279 \\ 0.0073 & 0.7405 & -0.6721 & -99.9548 \\ 0.9990 & -0.0358 & -0.0285 & 13.4519 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix} \quad (26)$$

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

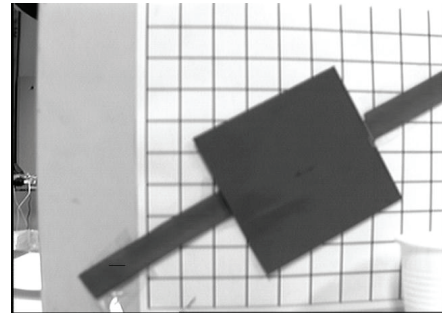
Table 1. Camera parameters

### 5.2. Verification of the visual measurement

An experiment was designed and conducted to verify the proposed method with a rectangular colour marker attached to a panel. A red rectangle was viewed as the valve, and had a dimension of 98mm × 100mm. Two green parts are used to simulate valve handles. The robot head was installed on the end of an industrial robot as shown in Fig.5 (a). The target was laid on the ground under the head. Images captured by two MINTRON 8055MK cameras in the head are as shown in Fig.5 (b).



(a) The experiment scene



(b) Image

Fig. 5. The experimental scene and target image

In the experiment, the target was fixed on the ground under the robot head. The position and pose of the robot's hand was changed so that the cameras could capture the fixed target. The position and pose of the target relative to the left camera at the  $i$ -th sampling is denoted as  $T_{ci}$ , that of the robot's hand as  $T_{ei}$ , and that of the target in the world frame of the industrial robot as  $T_{wi}$ . The edges of the red rectangle were detected using a Hough transformation (Tzvi 1990) after distortion corrections. Two points were then selected from each line to calculate,  $T_{ci}$ , the position and pose of the target according to the method described in Section 4. Four vertexes were computed using the intersection of these lines, then the fine position vector  $\bar{p}$  was obtained and  $T_{ci}$  was modified. Table 2 shows verification results.

### 5.3. A comparison with traditional stereovision

To compare the proposed method with a traditional stereo vision method, another experiment was conducted. Four points that the rectangle intersects with the x-axis and y-axis of the object frame were selected as feature points for stereovision. Their positions in Cartesian space were computed, and were used to

Times	$T_{ei}$ (robot end-effector)	$T_{ci}$ (visual measurement)	$T_{wi}$ (target in the world frame)
1	$\begin{bmatrix} 0.0162 & -0.1186 & 0.9928 & 1100.3 \\ 0.9269 & -0.3706 & -0.0594 & -275.3 \\ 0.3749 & 0.9212 & 0.1040 & 357.7 \end{bmatrix}$	$\begin{bmatrix} 0.9989 & 0.0195 & -0.0418 & 2.0 \\ -0.0362 & 0.8924 & -0.4497 & -9.0 \\ 0.0286 & 0.4508 & 0.8922 & 320.4 \end{bmatrix}$	$\begin{bmatrix} 0.9951 & -0.0827 & 0.0542 & 1142.1 \\ -0.0838 & -0.9963 & 0.0183 & -326.2 \\ 0.0524 & -0.0228 & -0.9984 & -6.9 \end{bmatrix}$
2	$\begin{bmatrix} -0.0530 & 0.1812 & 0.9820 & 1185.6 \\ 0.9200 & -0.3735 & 0.1186 & -282.7 \\ 0.3883 & 0.9097 & -0.1469 & 352.4 \end{bmatrix}$	$\begin{bmatrix} 0.9675 & -0.1499 & 0.2037 & -1.0 \\ 0.2261 & 0.8738 & -0.4306 & -10.1 \\ -0.1134 & 0.4627 & 0.8792 & 313.2 \end{bmatrix}$	$\begin{bmatrix} 0.9957 & -0.0811 & 0.0455 & 1141.2 \\ -0.0822 & -0.9964 & 0.0212 & -326.2 \\ 0.0436 & -0.0248 & -0.9987 & -6.2 \end{bmatrix}$
3	$\begin{bmatrix} 0.1149 & 0.2805 & 0.9530 & 1250.1 \\ 0.7659 & -0.6359 & 0.0948 & -394.9 \\ 0.6326 & 0.7190 & -0.2879 & 348.0 \end{bmatrix}$	$\begin{bmatrix} 0.9304 & -0.1362 & 0.3404 & 0.3 \\ 0.1848 & 0.9761 & -0.1144 & -12.6 \\ -0.3167 & 0.1694 & 0.9333 & 324.3 \end{bmatrix}$	$\begin{bmatrix} 0.9958 & -0.0860 & 0.0313 & 1141.9 \\ -0.0860 & -0.9963 & -0.0007 & -326.9 \\ 0.0313 & -0.0021 & -0.9995 & -4.8 \end{bmatrix}$
4	$\begin{bmatrix} -0.1446 & -0.0269 & 0.9891 & 1098.3 \\ 0.4550 & -0.8895 & 0.0423 & -531.1 \\ 0.8787 & 0.4562 & 0.1409 & 294.5 \end{bmatrix}$	$\begin{bmatrix} 0.9903 & -0.1097 & -0.0852 & 3.8 \\ 0.1290 & 0.9538 & 0.2712 & -11.0 \\ 0.0516 & -0.2796 & 0.9587 & 311.9 \end{bmatrix}$	$\begin{bmatrix} 0.9952 & -0.0964 & 0.0195 & 1140.8 \\ -0.0961 & -0.9952 & -0.0140 & -328.1 \\ 0.0207 & 0.0121 & -0.9997 & -3.6 \end{bmatrix}$
5	$\begin{bmatrix} 0.1183 & 0.1075 & 0.9871 & 1198.4 \\ 0.3576 & -0.9320 & 0.0587 & -583.9 \\ 0.9263 & 0.3461 & -0.1488 & 294.5 \end{bmatrix}$	$\begin{bmatrix} 0.9717 & -0.1208 & 0.2030 & -0.3 \\ 0.0462 & 0.9400 & 0.3380 & -22.5 \\ -0.2317 & -0.3191 & 0.9190 & 343.2 \end{bmatrix}$	$\begin{bmatrix} 0.9957 & -0.0916 & 0.0146 & 1142.6 \\ -0.0919 & -0.9955 & 0.0230 & -328.3 \\ 0.0125 & -0.0242 & -0.9996 & -3.1 \end{bmatrix}$
6	$\begin{bmatrix} 0.1555 & -0.0686 & 0.9855 & 1142.4 \\ 0.9485 & -0.2683 & -0.1683 & -241.5 \\ 0.2760 & 0.9609 & 0.0233 & 352.5 \end{bmatrix}$	$\begin{bmatrix} 0.9910 & 0.1292 & 0.0363 & 1.7 \\ -0.0891 & 0.8357 & -0.5420 & -11.2 \\ -0.1004 & 0.5338 & 0.8396 & 318.3 \end{bmatrix}$	$\begin{bmatrix} 0.9948 & -0.0847 & 0.0571 & 1142.9 \\ -0.0853 & -0.9963 & 0.0079 & -327.0 \\ 0.0563 & -0.0127 & -0.9983 & -6.7 \end{bmatrix}$

Table 2. Verification Experiment Results

calculate the origin position, the X axis and Y direction vectors of the object frame. Thus the position and pose of the target relative to left camera was obtained.

Measurements were taken three times under the same conditions. Table 3 shows measuring results for the position and pose of an object. The first column shows the results for the position and pose of the object computed with a traditional stereovision method, while the 2nd column shows the results for the proposed method. Position values are shown in mm. The results with stereo vision were different, and the results with our method were unchanging.

Times	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 pose of an object using stereovision and rectangle constraint

Table 4 shows the positioning results for four feature points P1 to P4 in terms of the stereovision method and

our proposed method. The results with the proposed method are formed using the coordinates of the feature points in the object frame, and the position and pose of the object frame. It can be found that the positioning results with our method are very stable.

Times	Results with stereovision				Results with the proposed method			
	$P_1$	$P_2$	$P_3$	$P_4$	$P_1$	$P_2$	$P_3$	$P_4$
1	35.3	132.0	122.5	44.4	57.5	124.6	118.1	64.0
	47.0	80.6	-8.2	135.8	64.7	88.1	25.7	127.2
	948.7	899.8	948.0	888.3	977.9	941.3	976.8	942.4
2	34.6	131.1	121.6	45.1	57.5	124.6	118.1	64.0
	49.2	80.6	-8.1	135.5	64.7	88.1	25.7	127.2
	955.8	893.8	941.6	882.6	977.9	941.3	976.8	942.4
3	34.5	132.4	122.7	44.6	57.5	124.6	118.1	64.0
	47.4	80.5	-8.0	135.6	64.7	88.1	25.7	127.2
	957.1	899.3	948.2	888.5	977.9	941.3	976.8	942.4

Table 4. Positioning results for the feature points using stereovision and rectangle constraints

It should be noticed that the method proposed in this paper computes the position and pose of the target with the imaging points on edge lines of the rectangle (these are detected through a Hough transformation). Even if there were errors in some imaging points, the edge line would be accurate enough for the Hough transformation, which can eliminate the influence of random errors. Furthermore, it does not need to employ feature point matching. Measuring results for the proposed method are

stable and insensitive to random noise. In other words, the proposed method is more robust than the traditional stereovision method in terms of noise resistance.

Errors which exist in the measurements taken with the proposed method mainly occur due to system errors such as camera calibration errors and so on. Errors in poses should be less than those in positions, i.e. the object pose measurements have higher accuracy.

### 6. Experimental results

Based on the proposed method, experiments were designed and conducted for our humanoid robot to approach and operate a valve with a rectangular coloured marker attached on its panel, as shown in Fig. 2. The red rectangle was 100mm in height and 100mm in width. The pose of the valve handles was marked by green colour, whose direction was consistent with the X axis in Fig. 3. The head with two MINTRON 8055MK cameras is shown in Fig. 1. Two mini cameras were fixed on the wrists of the two manipulators. The cameras on the head were well calibrated, but the ones on the wrists are not calibrated.

#### 6.1 Approaching the Valve by the mobile base

At the beginning, the robot searched for the target valve in the laboratory. When the valve was found, the 1st stage described in Section 2 was started. When the valve was within two meters of the robot, the 2nd stage began and the method described above was in operation. The position and pose of the mobile base was adjusted according to that of the valve until the robot was in an adequate operational area. When the robot stopped moving, the position and pose of the valve relative to the head was measured again by using the proposed method. The position and pose of the target valve relative to the chest of the humanoid robot could be obtained through coordinate transformation. Table 5 shows the position and pose of the target relative to the reference frame at the chest. The pose and position of the target relative to two end-effectors could also be calculated respectively through coordinate transformation.

At the 1st and 2nd stages, both arms were not in operation and were kept in a static position and pose. In the process of approaching the target by the humanoid robot, two arms were positioned so that they did not block the head field of view of the target valve.

$\bar{n}$	$\bar{o}$	$\bar{a}$	$\bar{p}$ (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 5. Position and pose of the Valve

#### 6.2. Moving the end-effectors near to the Handle

Once the robot was in an adequate operational area, the hands of both arms would move, one to each of the handles of the valve. At the same time, the cameras on

the head were inactive. The goal position and pose of the two end-effectors were determined according to the pose and position of the valve given above. The goal positions of the hands, especially in the cameras' view direction, had an offset added in order to avoid collisions between the end-effectors and the valve as a result of any error. The moving paths were planned to satisfy positions with a high priority so as to avoid collisions, except for the cameras' view direction.

The movements were controlled using kinematic and inverse kinematic models of manipulators. Therefore, end-effectors could move to the given goal quickly. At the same time, the camera at each hand was in operation to measure the size of image areas of the valve handle (the green colour marker on both sides of the valve). The size of the green markers increases as each hand moved closer to the handle. If the size was large enough or a given position was reached, position adjustment was ended and the process changed to the 4th stage of the proposed positioning method.



(a) The left image



(b) The right image

Fig. 6. Images captured by the cameras on the robot head

Fig. 6 provides a pair of images of one hand, captured at the end of the 3rd stage. Both images, i.e. Fig. 6(a) and Fig. 6(b), were captured by the left and the right cameras of the robot head respectively. It can be seen that the end-effector is at the place near to the handle with an appropriate pose. This means that the pose calculated by the proposed method has good accuracy.

#### 6.3. Approaching and Catching the Handle

An image-based visual servoing method in the 4th stage was applied to guide the end-effectors to reach and catch each handle. As pointed out in (Hager 1996), image-based visual servoing methods for an eye-in-hand system

have the drawback that a target object may be out of the camera's field of view during pose adjustment of the end-effector, which results in servoing failure. If positions were only adjusted in a stationary pose, this drawback would be overcome. However, to ensure that the pose is stationary in the servoing process and the end-effector can catch the handle with appropriate pose, the pose of the end-effector should be given accurately at the beginning of visual servoing. This is why the pose of the valve needs to be measured accurately in the 3rd stage of the proposed positioning method and kept unchanged in the 4th stage.

The goal of image-based visual servoing is that the image of the green marker, representing the handle, should match a given reference image as much as possible. The position adjustments of end-effectors were given a high priority, except for one (in the cameras' view direction), to avoid collision with valve handles. The end-effector was open during the visual servoing process. The goal was to adjust the end-effector position at a small range, and the gripper reached the handle with an appropriate pose when guided by the camera in hand. The final part of the process involved the gripper closing to grasp a handle. A hybrid control method using force and position was employed to rotate the valve with the robot's two hands. It is omitted here.

In a series of experiments, the humanoid robot was able to autonomously find, reach and operate the valve successfully. These experiments show that the position and pose of the valve calculated using the proposed methods are accurate enough to guide two arms in order for them to operate the valve. The advantages of using both eye-to-hand and eye-in-hand systems are clearly demonstrated.

## 7. Conclusions

A new visual servoing strategy for a humanoid robot to approach and grasp a valve is proposed. It consists of four stages, namely rough base approaching, fine base approaching, rough hand approaching and fine hand approaching and grasping. As an important part in the process of autonomous valve manipulation, a visual positioning and control method was proposed in this paper (for a hand-eye system and rectangular shape constraints). It employs multiple imaging points, which lie on lines with pre-known parameters in the objective frame. Positioning accuracy and robustness, especially the pose, were increased, and the influence of position errors in images was eliminated.

Based on the position and pose of the valve being calculated using the proposed method, end-effectors could smoothly reach valve handles, under the guidance of a hand-eye system. The end-effectors of our humanoid robot could catch the handles successfully and rotate the valve. The results verify the effectiveness of our proposed methods. The reliability and robustness of the system were significantly improved. The methods employed

can be widely applied in real-world applications of humanoid robots and mobile manipulators.

**Acknowledgement:** The authors would like to thank the National High Technology Research and Development Program of China for supporting this work under grant 2002AA422160. We would also like to thank the National Key Fundamental Research and Development Project of China (973, No.2002CB312200) for their support.

## 8. References

- Bartoli, A., Sturm, P. & Horaud, R. (2001). Structure and motion from two uncalibrated views using points on planes, *Proceedings of the 3rd International Conference on 3D Digital Imaging and Modelling*, pp. 83-90.
- Cardenas, A.; Goodwine, B.; Skaar, S. & Seelinger, M. (2003). Vision-based control of a mobile base and on-board arm. *The International Journal of Robotics Research*, Vol. 22, No. 9, pp. 677-698
- Chaumette, F. & Malis, E. (2000). 2-1/2D visual servoing: a possible solution to improve image-based and position-based visual servoing, *Proceedings of IEEE International Conference on Robotics & Automation*, pp. 630-635
- Corke P. I. & Hutchinson, S. A. (2000). A new partitioned approach to image-based visual servo control, *Proceedings of the 31st International Symposium on Robotics*, Montreal.
- Flandin, G.; Chaumette, F. & Marchand, E. (2000). Eye-in-hand /eye-to-hand cooperation for visual servoing, *Proceedings of IEEE International Conference on Robotics and Automation*, San Francisco
- Hager, G. D.; Hutchinson, S. & Corke, P. I. (1996). A tutorial on visual servo control. *IEEE Transaction on Robotics and Automation*, Vol.12, No. 5, pp. 651-670
- Han, M.; Lee, S.; Park, S.-K. & Kim. M. (2002). A new landmark-based visual servoing with stereo camera for door opening, *International Conference on Control, Automation and Systems*, Muju Resort, Jeonbuk, Korea, pp.1892-1896
- Heikkela, T.; Sallinen, M.; Matsushita, T. & Tomita, F. (2000). Flexible hand-eye calibration for multi-camera systems, *Proce. of IEEE/RSJ International Conference on Intelligent Robots & Systems*, pp.2292-2297
- Tzvi, D. B. & Sandler, M. B. (1990). A combinatorial Hough transform. *Pattern Recognition Letters*, Vol. 11, No.3, pp.167-174,
- Wells, G. & Torras, C. (2001). Assessing image features for vision-based robot positioning. *Journal of Intelligent and Robotic Systems*, Vol.30, pp.95-118
- 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
- Zhu, S. & Qiang, X. (2000). Analysis of 3-D coordinate vision measuring methods with feature points on workpiece. *Journal of Optics and Precision Engineering*, Vol.8, No.2, pp.192-197