Visual Servoing Control of Baxter Robot with Obstacle Avoidance using Kinematic Redundancy

C. Yang\textsuperscript{1,2}, S. Amarjyoti\textsuperscript{3}, X. Wang\textsuperscript{1,3}, Z. Li\textsuperscript{2}, H. Ma\textsuperscript{3} and C.-Y. Su\textsuperscript{2,4}

\textsuperscript{1}Center for Robotics and Neural Systems, Plymouth University, UK
\textsuperscript{2}College of Automation Science and Engineering, South China University of Technology, Guangzhou, China.
\textsuperscript{3}School of Automation, Beijing Institute of Technology, Beijing, China
\textsuperscript{4}Department of Mechanical and Industrial Engineering, Concordia University, Canada

Abstract. In this paper, a visual servoing control enhanced by an obstacle avoidance strategy using kinematics redundancy has been developed and tested on Baxter robot. A Point Grey Bumblebee2 stereo camera is used to obtain the 3D point cloud of a target object, which is then utilized to manipulate the closely coupled dual arm Baxter robot. The object tracking task allocation between two arms has been developed by identifying workspaces of dual arms and tracing the object location in a convex hull of the workspace. By employment of a simulated artificial robot as a parallel system as well as a task switching weighting factor, the actual robot is able to restore back to the natural pose smoothly in the absence of the obstacle. Two sets of experiments have been carried out to demonstrate the effectiveness of the developed servoing control method.

Keywords: Visual servoing, Point cloud, Obstacle Avoidance

1 Introduction

Visual sense is one of the most essential markers of intelligence and contributes immensely to the interactions with our environment. Visual feedback plays an important role for an anthropomorphic robot to operate in human surroundings. Especially, the advancements in 3D Vision Guided Robots (VGRs) have led to accurate non-contact geometrical measurements and reduction of workspace ambiguity. Therefore, strides made in robotic and machine vision are of paramount importance for engineering, manufacturing and design processes [1]. Visual servoing (VS) is defined as the use of visual feedback mechanisms for the kinematic control of a robot. Based on the positioning of the camera on the link and control techniques, VS ramifies into several types. Eye-in-hand and eye-to-hand VS are represented by the position of the camera on the robotic manipulator. Being attached on the robot arm, eye-in-hand VS provides a narrower field of view as compared to eye-to-hand servoing. Many control schemes use either a direct visual servoing or a dual loop system [6]. A visual processing method for hand gesture recognition is developed in [8] for control a simulated iCub robot.
on YARP platform. In [7], a noncontacting vision-based method of robot tele-operation is presented, whereas a human operator communicate simultaneous six-degree-of-freedom motion tasks to a robot manipulator by having the operator perform the three-dimensional human hand-arm motion that would naturally be used to complete an object manipulation task.

These schemes are robust but come at the price of reduced performance and speed variables. The drawbacks of time delay between image acquisition and its relation to latency in robot controller can be solved by the use of an open loop based VS for computing the reference input of the joint controller from the targets projection once [3]. It is practically useful to integrate visual sensing with other sensing technologies to develop a hybrid sensing system for better control performance. A combination of force sensor, CCD cameras and laser projector is used for accurate drilling on a 3D surface of unknown pose [2]. Eye-to-hand pose estimation and eye-in-hand object grasping in a hybrid control loop have been studied in [11][9][13]. In this paper, we use a similar approach that subsumes stereo image processing for pose estimation and the range data from IR sensors near the end effectors for precise gripping.

Similar to our human arm, the Baxter robot manipulator is of 7 joints. When the degree of freedom (DOF) in joint space is larger that that required in the task space, there is redundancy. At most 6DOF is required for manipulation in the Cartesian space, and the redundancy can be used to generate motion for obstacle avoidance. In [10, 14], the decomposition of the solution into a particular and a homogeneous component effectively illustrates the priority of the multiple goals that is exact end-effector control with redundant degrees of freedom maximizing the distance to obstacles. In this work, the robot end effector is controlled to follow the reference trajectory generated according to the detected target object’s position, while at the same time the joint space redundancy is utilized to avoid obstacle in the environment. Using our proposed method, the redundant Baxter robot arm will avoid obstacle without sacrificing the tracking performance of the end effector. In addition, the develop control method also enables the robot arm to restore its natural pose when the obstacle is absent, by employment of a simulated artificial robot as a parallel system based on kinematic model of the robot.

2 Baxter® Robot and its Kinematics

2.1 Workspace identification of the Baxter® robot arms

Baxter® robot is a semi-humanoid robot consisting of two arms and a rotational head on its torso, which can be installed on a movable pedestal. Several attributes such as the presence of a CCD camera and an IR range sensor at the end effector as shown in Fig. 1(a), make Baxter the definable option. As illustrated in Fig. 1(b), each arm of Baxter robot constitutes of 8 links and 7 rotational joints with different interchangeable grippers, e.g., electric gripper or a vacuum cup. The kinematics model of a Baxter robot arm was built in our previous work [5], where the details of DH parameters as well as joint rotation
limits are provided. Robot manipulator workspace boundary estimation is essential for improvising of robotic algorithms and optimization of its overall design and analysis. In this paper, we first extend the Monte Carlo method used on a single arm [5] to dual arms of Baxter robot, in order to identify the reachable workspace of two arms. Homogenous radial distribution is used to generate 6000 points in the joint space for each arm separately. The joint angle values are chosen randomly and forward kinematics is implemented to evaluate the end effector points, creating a point cloud of the reachable workspace for dual arms. The generated point cloud of the workspace for both the manipulators is shown in Fig. 2(a). Next, Delaunay triangulation is applied to the points in the 3D space to generate a set of points with a circumcircle without any points in its interior. This facilitates the creation of a convex hull of the joint space. It is used to constrain the workspace and to distinguish the ranges of control between the left and right arms for efficient manoeuvrability.
3 Visual Sensor and its Preprocessing

A client-server UDP network as shown in Fig. 3(a) is employed to integrate various system components and confers parallel processing. The machine vision processing is done in the client computer connected with a Point Grey® Bumblebee2 stereo camera with IEEE-1394 Firewire connection. The Point Grey Bumblebee2 stereo camera, as shown in Fig. 3(b), is a 2 sensor progressive scan CCD camera with fixed alignment between the sensors. Video is captured at a rate of 20 fps with a resolution of 640x480 to produce dense colored depth maps to assist in tracking and a viable pose estimation of the object. The resolution-speed trade-off has to be managed carefully as an increased frame speed gives a smooth robot trajectory whereas enhances the processing time. And an increased resolution provides a denser, more accurate point cloud for feature extraction but with increased latency.

![System architecture and Bumblebee2 camera](image)

**Fig. 3.** Left: Point Grey® Bumblebee2 camera; Right: communication network

### 3.1 Camera calibration

Camera calibration is necessary as the use of lenses introduces nonlinearities and deviates from the simple pin-hole model such as lens distortion, namely radial and tangential distortion. The camera parameters, namely the intrinsic, extrinsic and distortion are evaluated by the use of a 2D checker-board pattern. 3D reference models were avoided due to computational complexity and high cost of precise calibration objects. In this work, 20 checkerboard images were fed to the calibrator algorithm encompassing differential angles in the projection space. This provides enough values to estimate the camera geometry parameters, including 4 intrinsic parameters of \( f_x, f_y, c_x, c_y \), 5 distortion parameters of \( k_1, k_2, k_3 \) (radial) plus \( p_1, p_2 \) (tangential) and the camera extrinsic parameters of \( \Phi, \Psi, \Theta \) (rotation) together with \( T_x, T_y, T_z \) (translation). The point “Q” on the object plane is related with the image plane point “q” by the following equations.

A summary to the definition of these parameters is provided in Table 1.

\[
\begin{bmatrix}
    x \\
    y \\
    1 
\end{bmatrix} = sA[r_1 \ r_2 \ r_3 \ T] \begin{bmatrix} X \\ Y \\ 0 \end{bmatrix}, \quad x = f \frac{X}{Z} + c_x, \quad y = f \frac{Y}{Z} + c_y, \quad q = sHQ \quad (1)
\]
The camera parameters are finally obtained using Closed Form solution and Maximum Likelihood estimation. Further the distortion is nullified from the image points by introducing the following equations

\[ x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \quad y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \]

Table 1. Definition of variables

| X, Y | World co-ordinates of the point | \( e_x, e_y \) | Projection displacement parameter | H, A | Homography matrix, and Intrinsic matrix | f | Focal length of the camera lens | r_1, r_2, r_3 | spatial rotational matrices | s | scaling factor | T | translational matrix |

The co-ordinate transformation of the detected feature points from the Bumblebee2 co-ordinates to Baxter co-ordinates can be achieved by a homogeneous transformation matrix to be obtained as

\[ T = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 & X_2 & X_3 & X_4 \\ Y_1 & Y_2 & Y_3 & Y_4 \\ Z_1 & Z_2 & Z_3 & Z_4 \\ 1 & 1 & 1 & 1 \end{bmatrix}^{-1} \]

where \((x_i, y_i, z_i)\) and \((X_i, Y_i, Z_i), i = 1, 2, 3, 4\) are four non-collinear points measured from the robot coordinate and the Bumblebee2 coordinate, respectively.

4 Object Detection and Localization

4.1 3D reconstruction

![Images captured from Bumblebee2](image)

Fig. 4. Images captured from Bumblebee2

The images captured by the Bumblebee2 stereo camera in active ambient lighting are shown in Fig. 4. Both the images are calibrated using the camera
intrinsics and are corrected for distortion. Subsequently, the undistorted images are stereo rectified in order to align the epipolar lines of both the projection planes and ensure the presence of similar pixels in a specified row of the image. The images obtained are then frontal parallel and are ready for correspondence estimate. The Essential and the Fundamental matrix are calculated by using Epipolar geometry. Essential matrix is a $3 \times 3$ matrix with 5 parameters; two for translation and three for the rotation values between the camera projection planes. The Fundamental matrix on the other hand, represents the pixel relations between the two images and has seven parameters, two for each epipole and three for homography that relates the two image planes. Bouguets algorithm is then implemented to align the epipolar lines and shift the epipoles to infinity. Fig. 5(a) depicts the results of stereo rectification where the red and cyan colors represent the left and right images with row aligned pixels.

![Fig. 5. Stereo images and 3D reconstruction.](image)

Table 2. Definition of variables

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$x^l$</td>
<td>column value of left image pixel</td>
</tr>
<tr>
<td>2</td>
<td>$x^r$</td>
<td>column value of right image pixel</td>
</tr>
<tr>
<td>3</td>
<td>$D$</td>
<td>Depth</td>
</tr>
<tr>
<td>4</td>
<td>$T$</td>
<td>Baseline</td>
</tr>
<tr>
<td>5</td>
<td>$f$</td>
<td>focal length</td>
</tr>
<tr>
<td>6</td>
<td>$d$</td>
<td>disparity</td>
</tr>
<tr>
<td>7</td>
<td>$Q$</td>
<td>Projection matrix</td>
</tr>
<tr>
<td>8</td>
<td>$X/W, Y/W, Z/W$</td>
<td>3D world co-ordinates</td>
</tr>
</tbody>
</table>

The definition of variables used below is provided in Table 2. Stereo correspondence is a method of matching pixels with similar texture across two co-planar image planes. The distance between the columns of these perfectly matched pixels is defined below:

$$d = x^l - x^r$$

Block matching is implemented for evaluating the correspondence between the images. Block sizes of 15 pixel window are used to find the matches by the use of SAD (sum of absolute differences). The disparity range is kept low as $[0, 40]$ in order to match the indoor low texture difference and taking into account computational speed. Semi Global method is used to force the disparity values to the neighboring pixels for a more comprehensive result. The disparity output
is shown in Fig. 5(b). Disparity is inversely proportional to the depth of the pixel and is related by the Triangulation equation (4). Triangulation refers to the estimation of depth of an object by visualizing its location from two different known points.

\[ D = \frac{T_f}{d} \] (4)

The reconstruction of the image in the Cartesian co-ordinates is obtained by the use of projection matrix evaluated using Bouguets algorithm (5). Fig. 5(c) depicts the 3D reconstruction of the robots workspace.

\[ Q[x, y, d, 1]^T = [X, Y, Z, W]^T \] (5)

### 4.2 Object detection

Color based segmentation is used in order to separate a single color object from the background. The image is converted into L*a*b* color space and the Euclidean distance between red-green and yellow-blue opponent components of the object and a and b matrices calculated. The minimum value gives the most accurate estimate of the object. Further, the corners of the object are calculated by Harris corner detector and the centroid calculated by intersection of the diagonals. The depth value of the centroid is then extracted from reconstructed point cloud of the task space. Fig. 6 shows the calculated robot co-ordinates of the centroid of the object after co-ordinate transformation.

**Fig. 6.** Object detection results

### 4.3 Switching scheme for dual arms

The depth of the object to be manipulated is mapped to the convex hull to expedite the process of decision making, because a point cloud matching would hinder the processing speed. This is done by checking the presence of the point’s respective co-ordinates in the convex-hull projection on the three Cartesian planes. Hence, if \((X_1, Y_1, Z_1)\) be the point representing the object, its presence in the 3D hull is detected by following five steps:

The above procedure ensures an efficient methodology to reach the object in the robot workspace. It mimics the human intuition of using the nearest
Algorithm 1 Object detection in the 3D workspace hull

1: Check if the XY plane projection of the hull contains the point \((X_1, Y_1)\), YZ plane contains point \((Y_1, Z_1)\) and XZ plane contains point \((X_1, Z_1)\) using Ray Casting Algorithm;
2: Obtain the presence decision in the workspace of both arms;
3: If the point is present in both of the manipulator workspace, give the control priority to the arm with the smallest Euclidean distance from manipulator;
4: Divide the control based on the detection of the point in left or right workspaces;
5: If the point lies outside both the workspaces, stop arm movement to avoid singularity.

possible arm for grabbing in order to avoid the use of excessive body movements and minimize the use of energy.

4.4 Detection of collision points

The collision points, \(p_{cr}\) and \(p_{co}\), are the two points either on the robot or on the obstacle, which covers the nearest distance between the robot and the obstacle. The forward kinematic of the robot is used to estimate the collision points. Each link of the robot can be seen as a segment in 3-D space. The coordinates of the endpoints of the segments, i.e. the coordinates of the joints, in Cartesian space can be obtained by forward kinematics.

The problem of estimating the collision points can be seen as the problem of searching the nearest point between the obstacle segment \(q_{l1} - q_{l2}\) and the segments that stand for the robot links. Firstly, the distance between segment \(q_{l1} - q_{l2}\) and the \(i^{th}\) segment of the robot links \([x_i, y_i, z_i] - [x_{i+1}, y_{i+1}, z_{i+1}]\) can be calculated by 3-D geometry, which is denoted by \(d_i\); the nearest points on the robot link and the obstacle are denoted by \(p_{cr_i}\) and \(p_{co_i}\), respectively. Then we have the collision point \(p_{cr} = p_{cr_{i_{\text{min}}}}\) and \(p_{co} = p_{co_{i_{\text{min}}}}\) and the distance \(d = d_{i_{\text{min}}}\) where \(i_{\text{min}} = \arg \min_{i=0,1,...,n} d_i\).

5 Obstacle Avoidance and Restoring Control

Once the target object’s co-ordinates has been detected, end effector desired trajectory for the visual tracking task can be defined as the moving trajectory of the target. While for visual guided manipulation task, the end effector desired trajectory can be generated according to the location of the target object. Denote the joint velocities as \(\dot{\theta}\) and end-effector velocity as \(\dot{x}\), then they must satisfy \(\dot{x} = J\dot{\theta}\) where \(J\) is the Jacobian matrix of each arm. If the dimension of \(\dot{\theta}\) is larger than the dimension of \(\dot{x}\), the manipulator will become kinematically redundant, which can be used to achieve some secondary goals, e.g. obstacle avoidance, in addition to the end-effector trajectory following task. For the inverse kinematic problem of the kinematically redundant manipulator, infinite number of solutions can be find. The general solution is given by (6).

\[
\dot{\theta} = J^\top \dot{x} + (I - J^\top J)z, \tag{6}
\]
where \( J^\dagger = J^T (JJ^T)^{-1} \) is the pseudo-inverse of \( J \) and \( z \) is an arbitrary vector, which can be used to achieve the obstacle avoidance [10].

5.1 Collision avoidance strategy

When the collision points \( p_{cr} \) and \( p_{co} \) are found, the manipulator needs to move away from the obstacle. The desired velocity \( \dot{x}_o \) moving away from the obstacle is designed as (7).

\[
\dot{x}_o = \begin{cases} 
0, & d \geq d_o \\
\frac{d_o - d}{d_o - d_c}v_{max} \frac{p_{cr} - p_{co}}{d}, & d_c < d < d_o \\
\frac{v_{max} p_{cr} - p_{co}}{d}, & d \leq d_c 
\end{cases}
\]

(7)

where \( d = \|p_{cr} - p_{co}\| \) is the distance between the obstacle and the manipulator; \( v_{max} \) is the maximum obstacle avoidance velocity; \( d_o \) is the distance threshold that the manipulator starts to avoid the obstacle; \( d_c \) is the minimum acceptable distance and the manipulator will avoid at the maximum speed.

On the other hand, in order to eliminate the influence of the obstructing when the obstacle has been removed, the manipulator is expected to restore its original state. To achieve that, an artificial parallel system of the manipulator is designed in the controller in real time to simulate its pose without the influence of the obstacle, as shown in Fig. 7, where the dashed black line indicates the parallel system. The restoring velocity \( \dot{x}_r \) is designed as a closed-loop system as

\[
\dot{x}_r = K_r e_r,
\]

(8)

where \( K_r \) is a symmetric positive definite matrix and \( e_r = [e_{r1} \ e_{r2} \ e_{r3}]^T \) is the position errors of the joints between the parallel system and the real system.

Based on our previous work [12], the control strategy is given below

\[
\dot{\theta}_d = J_e^\dagger (\dot{x}_d + K_e \dot{e}_x) + (I - J_e^\dagger J_e) [\alpha z_o + (1 - \alpha)z_r] 
\]

(9)

where

\[
\begin{align*}
z_o &= [\dot{x}_o^T J_o (I - J_e^\dagger J_e)]^\dagger (\dot{x}_o^T \dot{x}_o - \dot{x}_o^T J_o J_e^\dagger \dot{x}_e) \\
z_r &= [\dot{x}_r^T J_r (I - J_e^\dagger J_e)]^\dagger (\dot{x}_r^T \dot{x}_r - \dot{x}_r^T J_r J_e^\dagger \dot{x}_e)
\end{align*}
\]

(10)

5.2 Object gripping with hybrid servoing

Stereo camera provides a wider angle of view serving the eye to hand servoing system, while for meticulous sensing of the object, an IR range sensor equipped at the end effector provides an eye-in-hand feedback, and helps to reduces the processing time and provides satisfactory accuracy. The IR range sensor used provides 16 bit digital data and checks for the location of object along the Z
The artificial parallel system built using Baxter® kinematics model [5]. The solid black line indicates the real manipulator. The dashed black line indicates the manipulator in the artificial parallel system.

6 Experiment Studies

In order to demonstrate the proposed visual servoing method, two groups of experiments of visual tracking task and visual guided manipulation task are performed, with obstacle and without obstacle for comparison.

6.1 Visual tracking task

In the visual tracking experiments, the robot is controlled to follow the object using two manipulators. The experiment video frames are shown in Fig. 8 and Fig. 9 for the experiments with and without obstacle respectively. The end-effector trajectory of this two comparative experiments are shown in Fig. 10.

As can be seen, the robot automatically choose the appropriate manipulator to track the object based on the workspaces. When the obstacle is found near the manipulator, the corresponding manipulator change its pose to avoid collision and, at the same time, keep its end-effector position tracking the object.

6.2 Visual guided manipulation task

In the visual guided manipulation experiments, the robot is controlled to pick up the object and put it into a box at a predefined position. The end-effector trajectory is planed using the Algorithm 2.

The experiment video frames are shown in Fig. 11 and Fig. 12 for the experiments with and without obstacle respectively. The end-effector trajectory
Visual Servoing Control of Baxter Robot

Fig. 8. Video frames of visual tracking without obstacle without obstacle.

Fig. 9. Video frames of visual tracking with obstacle with obstacle.

Fig. 10. End-effector trajectory of visual tracking. The blue and red dashed lines indicates the reference end-effector trajectory of the left and right manipulators respectively; the blue and red solid lines indicates the actual trajectory of the end-effectors of the left and right manipulators respectively.

of this two comparative experiments are shown in Fig. 13. As can be seen, the robot can pick up the object precisely even if the position of object has been changed. When the obstacle is found near the manipulator, the corresponding manipulator change its pose to avoid collision but can still move the end-effector right above the box and drop the object into the box.

7 Conclusion

This paper developed and implemented a visual servoing control method with automatic obstacle avoidance on the dual arm Baxter robot. Stereo imaging and range sensors are utilized to attain eye-to-hand and eye-in-hand servoing, such that both target object and obstacle are detected and localized. Kinematic redundancy has been exploited for obstacle avoidance. A simulated artificial robot
**Algorithm 2** Visual guided manipulation trajectory planning

1: Move to the position 10cm behind the object;
2: Move forward step by step with a step length of 0.5cm;
3: Check the IR range sensor after each 0.5cm step;
4: If the range data is less than a threshold, grip the object and move up by 20cm;
5: Move to the predefined position above the box and release the object;

![Video frames of visual guided manipulation without obstacle.](image)

**Fig. 11.** Video frames of visual guided manipulation without obstacle.

![Video frames of visual guided manipulation with obstacle.](image)

**Fig. 12.** Video frames of visual guided manipulation with obstacle.

![End-effector trajectory of the visual guided manipulation.](image)

**Fig. 13.** End-effector trajectory of the visual guided manipulation. The dashed lines indicates the reference end-effector trajectory; and the solid lines indicates the actual trajectory of the end-effectors.

of same kinematics as Baxter robot is employed as a parallel system to enable the robot restoring back to the natural pose when the obstacle is absent. Two switching mechanisms have been designed, one for task allocation between two arms for visual tracking task, while the other for smooth switching between obstacle avoiding task and restoring task. An object tracking and an object manipulation experiments have been performed to test the develop control method, and
the results have demonstrated the effectiveness and robustness of the developed approach.

References