Multiple cameras visual servoing used for large scale 3D positioning

Yimin Zhao¹, Wen-Fang Xie², Member, IEEE, Xiao-Wei Tu³

¹) Department of Mechanical & Industrial Engineering, Concordia University, Montreal, QC, Canada, H3G 1M8 (E-mail: ymnzhao@yahoo.com).
²) Department of Mechanical & Industrial Engineering, Concordia University, Montreal, QC, Canada, H3G 1M8 (Tel: 514-848-2424, ext. 4193, Fax: 514-848-3175 E-mail: wfxtie@encs.concordia.ca).
³) National Research Council, NRC-AMTC, 5145 Av. Decelles, Montreal, QC, Canada, H3T 2B2 (E-mail: Xiao-Wei.Tu@cnrc-nrc.gc.ca)

Abstract—In this paper, a multiple cameras visual servoing system is presented to correct the pose of robot end-effector in large scale 3D positioning in robot manufacturing system. A 3D CAD laser projector is used to provide four target points on the projection screens for four cameras respectively. Eye-in-hand architecture is used to perform visual servoing and the image based visual servoing (IBVS) is adapted as the control strategy. This paper is focused on detecting the four target points by four cameras respectively. Theoretically, the image Jacobian matrix doesn’t hold in this configuration because the target points are floating in base frame along the laser beam during the visual servoing process. But in a step of iteration, the target points on the projection screens can be referred as fixed points which are also called local desired target points. The image features of local desired target points in image planes of cameras are the desired features in this iteration. By observing the four target points on projection screens, the IBVS controller can drive the end-effector of robot to its local desired position in each step of iteration and ultimately to its calibration position. In this configuration, the simple constant depths are adopted for obtaining the image Jacobian matrix during visual servoing. A simulation was carried out to verify the effectiveness of the proposed method.

Keywords—IBVS, Eye-in-hand, Multiple Cameras, Laser projector

I. INTRODUCTION

Visual servoing is a method in which visual feedback is introduced in the robot control loop to increase the accuracy of the overall robot system. This saves the need to increase the accuracy of different parts like end-effector and sensors attached to the system. In addition, visual servoing helps in controlling the robot pose with respect to the target even in the presence of calibration error [1].

In aerospace manufacturing industry, there is an increasing demand for automated robotic manufacturing system and the high precision is needed for large scale positioning, which is hard for traditional methods to obtain. Although visual servoing has been widely used in robotic manufacturing system, the size of the workpiece is generally large in aerospace industry and the target features should be chosen far away each other to achieve visual servoing. Hence it is hard to put all the target features inside the field of view (FOV) of single camera. Moreover, because work piece has a relative large size, the disposal of cameras is a key problem [2]. And sometimes, the target features are occluded from FOV of camera [3]. Unsuitable viewpoint may possibly cause a part of the object or some target feature points to get out of the image plane, which will make the robot control system become unstable. In addition, incorrect estimation of depth of the considered target features may cause convergence and stability problems. Some papers have proposed multiple cameras configuration to improve the visibility of object, such as stereo camera [4], multiple cameras [5], and two cameras- one is fixed on the end-effector, the other is fixed in the workspace [6]. But all these methods only increase the number of cameras and to give different views to observe the object. These methods are only effective in the case where the target object is relative small in size and the robot end-effector moves a long distance. In paper [7], a laser pointer was used to improve the accuracy of the depth estimation. In large scale work piece robot manufacturing, visual servoing is used to correct the pose of robot and this is not the same as the case mentioned above.

Although some commercial 3D metrology and tracking systems including Laser Tracker System (LTS), Laser Radar (LR), indoor Global Positioning System (iGPS), [8,9,10] and other optic-electronic positioning system are used for positioning and tracking in manufacturing and assembly, these systems are largely limited to speed or highly expensive [11,12].

In this paper, a new multi-image based visual servoing system was proposed. It has been shown that the minimum number of points to have the complete velocity screw of the robot end-effector uniquely defined through pseudo inverse image Jacobian matrix is four [13]. So the proposed system consists of four aligned cameras [14] which capture the images of a laser projector pattern to obtain 6 Degree of Freedom (DOF) information about a workpiece with large size. The system configuration is shown in Fig. 1. A simulation has been carried out to demonstrate the proposed servoing system can perform high precision 3D positioning task in large scale workpiece robot manufacturing system.

The organization of this paper is as follows. In section II, the system configuration is given and the coordinate frame is defined. In section III, the image-based visual servoing principle is reviewed and the control law is derived in case of multiple cameras' configuration. In section IV, the simulation results are given to show the effectiveness of proposed method. The error analysis is carried out in this section. The conclusion is given in section V.
II. PROBLEM STATEMENT

A. System configuration

In this paper, a precise robotic 3D positioning system is proposed by using four digital cameras, a 3D CAD laser projector, four projection screens and a 6-DOF robot.

The system uses Eye-in-hand architecture to perform visual servoing shown in Fig. 1. In this system, four cameras are installed on a frame fixed on the robot end-effector and each of cameras is used to look at one target feature on the projection screen. A 3D CAD laser which projects computer images (typically from CAD file) directly on objects for layout and alignment applications [15] is used to provide 4 target features on the projection screens at different file positions. It is noticed that at one working position, the laser beams depths of target features (laser spots on the projection screens) keep constant. The objective of the research is to drive the workpiece to a desired 3D position.

B. Coordinate frame definition

In Fig. 1, the relative frames and transformation matrices are defined as follow: \{B\} denotes robot base frame, \{E\} denotes the robot end-effector frame, and \{C_i\} denotes the frame of camera \(i\) \((i = 1,2,3,4)\). \(B^E T\) is the transformation matrix from \{E\} to \{B\}. \(C_i^E T\) is the transformation matrix from \{C_i\} \((i = 1,2,3,4)\) to \{E\}.

III. VISUAL SERVOING CONTROL LAW OF SYSTEM

A. Desired image feature coordinates

After calibration, \(p_1(X_1,Y_1,Z_1)\) is the coordinate of laser projector in robot base frame \{B\} and \(p_{01}(X_{01},Y_{01},Z_{01})\), \(p_{02}(X_{02},Y_{02},Z_{02})\), \(p_{03}(X_{03},Y_{03},Z_{03})\) and \(p_{04}(X_{04},Y_{04},Z_{04})\) are the coordinates in \{B\} of the intersection points of laser beams and projection screens. The image features of \(p_{01}\), \(p_{02}\), \(p_{03}\) and \(p_{04}\) in corresponding cameras’ image plane are denoted as \(f_{01}\).

B. Local desired image feature coordinates in an iteration

If the end-effector has an error from the desired position, the target points on the projection screens will be the intersection points of laser beams and the projection screens, which are denoted as \(p_i\) \((i=1,2,3,4)\) respectively. The parametric equation of \(i^{th}\) laser beam after calibration is denoted in robot base frame \{B\} as follows:

\[
X_i = X_{0i} t_i + X_i (1-t_i) \\
Y_i = Y_{0i} t_i + Y_i (1-t_i) \\
Z_i = Z_{0i} t_i + Z_i (1-t_i)
\]

where \(X_i, Y_i, Z_i\) are the coordinates of the point along the \(i^{th}\) laser beam and \(t_i\) is the ratio between distance from \(p_i\) to \(p_{0i}\) and distance from \(p_{0i}\) to \(p_{00}\).

In an iteration of visual servoing, local desired image feature \(f_{il}(x_{i},y_{i})\) \((i=1,2,3,4)\) is the projection of \(p_i\) in the image plane of the camera \(i\) \((i=1,2,3,4)\) which is at calibration position. To demonstrate the process of the computation of local desired image features, the camera 1 is taken as an example, which is shown in Fig. 2. The parametric equation of laser beam \(p_1-p_{01}\) is rewritten as:

\[
X_1 = X_{01} t_i + X_i (1-t_i) \\
Y_1 = Y_{01} t_i + Y_i (1-t_i) \\
Z_1 = Z_{01} t_i + Z_i (1-t_i)
\]

If \(C_i^B T\) is the transformation matrix of camera \(i\) at current position frame \{C_i\} in base frame \{B\} and \(C_{01}^B T\) is the transformation matrix of camera \(i\) in calibration position frame \{C_{01}\} in base
Similarly (4), we have 
\[
\begin{pmatrix}
X_{p1} \\
Y_{p1} \\
Z_{p1}
\end{pmatrix} = \begin{pmatrix}
X_{01} \\
Y_{01} \\
Z_{01}
\end{pmatrix} + \begin{pmatrix}
t_{1}X_{01} \\
t_{1}Y_{01} \\
t_{1}Z_{01}
\end{pmatrix}
\]
where \( t_{1} \) is the ratio between distance from \( p_{1} \) to \( p_{01} \) and distance from \( p_{1} \) to \( p_{01} \). 

By solving the equation (5), we get \( \mathbf{T} = \mathbf{T}_{C1} \), 
where \( \mathbf{T}_{C1} \) is the ratio between distance from \( p_{1} \) to \( p_{01} \) and distance from \( p_{1} \) to \( p_{01} \). So, the coordinates of intersection point \( p_{1} \) in base frame are as follows:

\[
\begin{pmatrix}
X_{p1} \\
Y_{p1} \\
Z_{p1}
\end{pmatrix} = \begin{pmatrix}
X_{01} \\
Y_{01} \\
Z_{01}
\end{pmatrix} + \begin{pmatrix}
t_{1}X_{01} \\
t_{1}Y_{01} \\
t_{1}Z_{01}
\end{pmatrix}
\]

The coordinates of intersection point \( p_{1} \) in camera 1 calibration position frame \( \mathcal{C}_{01} \) are as follows:

\[
\begin{pmatrix}
X_{d1} \\
Y_{d1} \\
Z_{d1}
\end{pmatrix} = \begin{pmatrix}
\lambda X_{p1} \\
\lambda Y_{p1} \\
\lambda Z_{p1}
\end{pmatrix}
\]

In image plane of camera \( I \), local desired image coordinate can be calculated as follows:

\[
\begin{pmatrix}
x_{d1} \\
y_{d1}
\end{pmatrix} = \begin{pmatrix}
\lambda X_{d1} \\
\lambda Y_{d1}
\end{pmatrix}
\]

Similarly

\[
\begin{pmatrix}
x_{d2} \\
y_{d2}
\end{pmatrix} = \begin{pmatrix}
\lambda X_{d2} \\
\lambda Y_{d2}
\end{pmatrix}
\]

\[
\begin{pmatrix}
x_{d3} \\
y_{d3}
\end{pmatrix} = \begin{pmatrix}
\lambda X_{d3} \\
\lambda Y_{d3}
\end{pmatrix}
\]

\[
\begin{pmatrix}
x_{d4} \\
y_{d4}
\end{pmatrix} = \begin{pmatrix}
\lambda X_{d4} \\
\lambda Y_{d4}
\end{pmatrix}
\]

The vector of the local desired image features (in iteration) is as follow:

\[
f_{d1} = \begin{pmatrix}
x_{d1} \\
y_{d1}
\end{pmatrix}, \quad f_{d2} = \begin{pmatrix}
x_{d2} \\
y_{d2}
\end{pmatrix}, \quad f_{d3} = \begin{pmatrix}
x_{d3} \\
y_{d3}
\end{pmatrix}, \quad f_{d4} = \begin{pmatrix}
x_{d4} \\
y_{d4}
\end{pmatrix}
\]

\[
f_{di} = \begin{pmatrix}
f_{d1} \\
f_{d2} \\
f_{d3} \\
f_{d4}
\end{pmatrix}
\]

C. Control law of multiple camera system

Let \( \mathbf{v}_{E} = [v_{x}, \ v_{y}, \ v_{z}, \ \omega_{x}, \ \omega_{y}, \ \omega_{z}]^{T} \) be velocity screw of the robot end-effector and \( \mathbf{v}_{C_{i}} = [v_{C_{ix}}, \ v_{C_{iy}}, \ v_{C_{iz}}, \ \omega_{C_{ix}}, \ \omega_{C_{iy}}, \ \omega_{C_{iz}}]^{T} \) be velocity screw of the camera \( i = 1, 2, 3, 4 \). The velocity transformation matrix from the frame of camera \( i \) to that of robot end-effector is denoted as follows:

\[
\mathbf{E}_{C_{i}} = \begin{pmatrix}
0 & E_{R} & E_{P}\end{pmatrix}
\]

where \( E_{R} \) is rotational component of transformation matrix \( E_{T} \) and \( E_{P} \) is the translational vector from the \( E \) to \( C_{i} \).

\[
\mathbf{v}_{E} = \begin{pmatrix}
E_{R} \mathbf{v}_{C_{i}} \\
0 \\
0
\end{pmatrix}
\]

\[
\mathbf{v}_{C_{i}} = \begin{pmatrix}
E_{R} \mathbf{v}_{E} \\
0 \\
0
\end{pmatrix}
\]

\[
f_{i} = \begin{pmatrix}
x_{i} \\
y_{i}
\end{pmatrix}, \quad (i = 1, 2, 3, 4)
\]

\[
\mathbf{f}_{i} = \begin{pmatrix}
x_{i} \\
y_{i}
\end{pmatrix}, \quad (i = 1, 2, 3, 4)
\]

are the image features.

\[
\mathbf{f}_{i} = \begin{pmatrix}
x_{i} \\
y_{i}
\end{pmatrix}, \quad (i = 1, 2, 3, 4)
\]

are the corresponding image feature velocities.

It is assumed that the effective sizes of a pixel \( (s_{x}, s_{y}) \) are constant to simplify the computation without loss of generality. The transformation between \( (x_{i}, y_{i})^{T} \) and the pixel indexes \( (u_{i}, v_{i})^{T} \) depends only on the intrinsic parameters.

In order to design the feedback control for robot based on the velocity of the feature points, we have the following relationship between the motion of image features and the physical motion the camera should hold.

\[
f_{i} = \mathbf{J}_{img}(f_{i}, Z_{i})\mathbf{v}_{C_{i}}
\]

For each feature point \( (x_{i}, y_{i}) \), the image Jacobian matrix is represented as follows:

\[
\mathbf{J}_{img}(f_{i}, Z_{i}) = \begin{pmatrix}
\lambda & 0 & -x_{i} & -y_{i} & \frac{\lambda^{2} + x_{i}^{2}}{\lambda} - y_{i} \\
0 & \frac{\lambda}{Z_{i}} & -y_{i} & -x_{i} & \frac{y_{i}^{2}}{\lambda}
x_{i} & y_{i} & \frac{\lambda}{Z_{i}} & & \end{pmatrix}
\]

Substitute the equation (8) into the Equation (9)
\[
\begin{align*}
\dot{f}_i &= J_{img}^i(f_i, Z_i) \begin{bmatrix}
\frac{E_R}{C_x} & sk(E_P c) & \frac{E_R}{C_y} \\
0 & \frac{E_R}{C_z} & 0
\end{bmatrix}^{-1} \begin{bmatrix}
v_F \\
\omega_E
\end{bmatrix} \\
&= E J_{img}^i(f_i, Z_i) \begin{bmatrix}
v_F \\
\omega_E
\end{bmatrix}
\end{align*}
\]

Let \( Z = [Z_1, Z_2, Z_3, Z_4]^T \) is the vector of the depths of feature points. \( f = [f_1^T, f_2^T, f_3^T, f_4^T]^T \) is the vector of image features and \( \dot{f} = [\dot{f}_1^T, \dot{f}_2^T, \dot{f}_3^T, \dot{f}_4^T]^T \) is the vector of image feature's velocity vector.

By stacking the image Jacobian matrix of feature point together, we obtain the overall Jacobian matrix as follows:

\[ J_{img}(f, Z) = \left[ J_{img}(f_1, Z_1), J_{img}(f_2, Z_2), J_{img}(f_3, Z_3), J_{img}(f_4, Z_4) \right]^T \]

Hence, the relationship between the motion of feature and end-effector velocity is:

\[ \dot{f} = J_{img}^*(f, Z) \dot{f} \]

The end-effector velocity is expressed as follows:

\[ \dot{r} = J_{img}^*(f, Z) \dot{f} \]

where \( J_{img}^*(f, Z) \) is the pseudo inverse of the image Jacobian matrix. If \( f_d \) is the vector of the desired image features, the error function is defined as \( e(f) = f - f_d \) and we impose \( \dot{e}(f) = -K e(f) \), a simple proportional control law is given by

\[ \dot{r} = -K J_{img}^*(f, Z) e(f) \]

where, \( \dot{r} \) is the end-effector’s velocity sent to robot controller, \( K \) is the proportional gain which tunes the exponential convergence rate toward \( f_d \).

D. Switching control of multiple camera system

The proposed algorithm is divided into two control stages which are as follows: 1) guiding the workpiece to nearby of the desired position step by step through driving the image features to the local desired image feature in each iteration. In this stage, \( f_d = f_{ad} \) (the vector of the local desired image features) 2) When workpiece’s position error between the current position and desired one is approaching zero, the vector of the local desired image features is approximately equal to that of the desired image features, i.e., \( f_d = f_{ad} \) (the vector of the desired image features). In this stage, the vector of the desired image features which is determined from calibration is used to compute \( e(f) \) i.e., the error of the image features. The control block diagram of IBVS with multiple cameras is shown in Fig. 3.

IV. SIMULATION RESULTS

In order to show the validity of the proposed scheme, some simulation results are shown in this section. As shown in Fig. 1, system is composed of a 6-DOF robot (Motoman UPJ), 4 cameras mounted on the robot end-effector and a 3D CAD laser projector which provides four aligned laser target points on the projection screens. The camera focal length \( \lambda \) is 6 (mm), the scaling factor

\[ \alpha_x = \alpha_y = 1.3513 (\text{pixels/mm}) \]

The coordinates (pixels) of desired feature points in four cameras’ image planes after calibration are (320, 240), (320, 240), (320, 240), (320, 240) respectively, which are the central points of image planes of cameras. The depth of feature points is set as 0.3 (m). The distance between cameras \( L=1 \) (m). Laser projector coordinates in base frame: \( p_4 = (0.92, 4.3) \)

The desired position of robot end-effector frame in base frame represented in transformation matrix is as follow:

\[ \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0.42 \\
0 & 0 & 1 & 0.3 \\
0 & 0 & 0 & 1 
\end{bmatrix} \]

The coordinates of target points (desired position shown in Fig. 4 and Fig. 7) in base frame can be represented as follows:

\[ p_{01}(0.5,0.42,0.3) , p_{02}(0.5,1.42,0.3) , p_{03}(-0.5,1.42,0.3) , p_{04}(-0.5,0.42,0.3) \]

Case 1.

The initial position of robot end-effector frame in robot base frame is shown in Fig. 4. The simulation results are shown as in Fig.4 to Fig. 6.

Fig. 4 The trajectories of points P1, P2, P3, P4 in base frame
Case 2.
The initial position of robot end-effector frame in robot base frame is shown in Fig. 7. The simulation results are shown as in Fig. 7 to Fig. 9.

Fig. 7 The trajectories of points P1, P2, P3, P4 in base frame

Fig. 5 Trajectories of image features in image planes

Fig. 6 Image feature errors in image plane
target points for visual servoing system at the different working positions of robot manufacturing systems. The control law for floating target points is proposed and successfully applied in simulation model of the system. The simulation results verify the effectiveness of the proposed method and also validate the feasibility of applying the multiple camera configurations to satisfy the need for high precision 3D positioning, where the size of work piece is relatively large in aerospace industry.

**REFERENCES**


