An Active Stereo Vision-Based Learning Approach for Robotic Tracking, Fixating and Grasping Control

Nan-Feng XIAO (1),(2)  Saeid Nahavandi (1)

(1) School of Eng. and Tech., Deakin University,
Geelong Victoria 3217, Australia
(2) Computer School, South China University of Technology,
Guangzhou, 510641, China
E-mail: xiao@deakin.edu.au

Abstract
In this paper, an active stereo vision-based learning approach is proposed for a robot to track, fixate and grasp an object in unknown environments. First, the functional mapping relationships between the joint angles of the active stereo vision system and the spatial representations of the object are derived and expressed in a three-dimensional workspace frame. Second, the self-adaptive resonance theory-based neural networks (ART_NN) and the feed-forward neural networks (FF_NN) are used to learn the mapping relationships in a self-organized way. Third, the approach is verified by simulation using the models of an active stereo vision system which is installed in the end-effector of a robot. Last, the simulation results confirm the effectiveness of the present approach.

Keywords: Robot Control, Active Stereo Vision, Neural Network, Tracking, Fixating, Grasping

1. Introduction
Vision-based robotic tracking, fixating and grasping control is very important for robot navigation, object recognition, visual servoing. At present, the vision-based robot control is mainly based on CCD cameras (1)-(4), the robot control only relay on the robotic kinematics and dynamics, and the robotic learning ability is not considered or used in the tracking, fixating and grasping. However, because the relationships of the spatial coordinates of an object and the joint angles of a robot with the CCD cameras are highly nonlinear, the coordinate transformations among the visual frames and joint frames are very complicated and difficult to solve in the existing vision-based robot control.

Vision-based robotic tracking, fixating and grasping control depends on many environmental factors in an unknown environment, the robot control systems lack robustness, and the calibration of the CCD cameras is very slow and tedious in the existing methods. Although the existing methods can solve some problems, it is necessary to rely on the time-consuming and complicated 3-D reconstruction algorithms (8)-(9). Therefore, it is necessary to develop a more effective vision-based robotic tracking, fixating and grasping method, and use the robotic learning ability to improve the tracking, fixating and grasping in the unknown environment.

This paper presents an active stereo vision-based learning approach for robotic tracking, fixating and grasping. First, the many-to-one functional mapping relationships are derived to describe the spatial representations of the object in the workspace frame. Then, ART_NN and FF_NN are used to learn the mapping relationships, so that the active stereo vision system guides the end-effector to track, fixate and grasp the object without the complicated coordinate transformations and calibration. Last, the present approach is verified by simulation.

2. Visual Tracking, Fixating and Grasping
Active vision can easily realize selective attention and prevent an object to go out of the view fields of the cameras, therefore the active stereo vision-based robotic tracking, fixating and grasping can achieve greater flexibility in an unknown environment.

Figure 1 shows an active stereo vision-based robotic system for the tracking, fixating and grasping. The CCD cameras have 5 DOF, the robot has 6 DOF, which constitute a 11 DOF tracking, fixating and grasping system.

Because the active CCD cameras and the robot can move independently or together, the active CCD cameras can observe freely an object in . According to the visual feedback information, the robot can track, fixate and grasp autonomously the object.
3. Many-to-One Mapping Relationships

Figure 2 shows the projective relationships between the active stereo vision system and the object in $\omega$. Let $p_i (i=1,2,3)$ be a feature point on the object. When the active stereo vision system tracks $p_i$ and its image coordinates are registered in the centers $(O_i, O)$ of the left and right image planes of the two CCD cameras, respectively, the active stereo vision system is known as the fixation on $p_i$.

Let $Q_i = [\theta_1, \theta_2, \ldots, \theta_n]^T$. $P_i = \begin{bmatrix} x_{z_i} \\ y_{z_i} \\ z_{z_i} \end{bmatrix}$ and $V_i = \begin{bmatrix} u_{z_i} \\ v_{z_i} \\ w_{z_i} \end{bmatrix}$ be a joint vector of the active stereo vision system for the $i$th point, a spatial representation vector of $p_i$ in $\omega$, and an image coordinate vector of $p_i$ on the left and right image planes, respectively. It is known from Fig.2 that when $p_1$ and $p_2$ are visible to the CCD cameras, $Q_2$ and $P_2$ of $p_2$ identified by the active stereo vision system tracking on $p_2$ should be different from $Q_1$ and $P_1$. If another joint angle vector $Q_3$, $P_3$ is obtained by tracking $p_3$, $p_1$ and $p_2$ are still visible, $Q_1$ and $P_1$ are not change from that obtained by tracking on $p_1$ and $p_2$, respectively, despite the image coordinate vectors $V_1$ and $V_2$ change on the image planes of the CCD cameras. Therefore, there exist many combinations of $Q_i$ and $V_i$ which correspond to the same $P_i$, which means that $P_i$ is invariant to the changing $Q_i$ and $V_i$.

According to the projective geometry, $V_i$ can be expressed as follows,

$V_i = \Phi_i (Q_i, P_i), \quad (i=1,2,3)$, (1)

where $\Phi_i$ is a nonlinear projective function which maps the object and the joint angles on the left and right image planes of the CCD cameras. Therefore, $P_i$ is specified as

$P_i = (V_i, Q_i), \quad (i=1,2,3)$, (2)

where is a nonlinear many-to-one mapping function, which denotes that the combinations of $Q_i$ and $V_i$ correspond to $P_i$. On the other hand, it is known from Fig.2 that any combination of $Q_i$ and $V_i$ should map to the same $P_i$, because $p_i$ is stationary feature.

Fig. 1 A robot system with active vision

Fig. 2 Projection and fixation relationships

4. Active Vision-Based Robot Control

4.1 Tracking and Fixating Control

It is known from Fig.2 that when the active stereo vision system tracks $p_i$, $V_i$ is obtained for $P_i$ and $Q_i$, we have from Eq.(2), $P_i = (V_i, Q_i)$, where the active stereo vision system tracks $p_i$. When the active stereo vision system fixes $p_i$ and the image coordinate vector $V_{Oi}$ of $p_i$ corresponds to the centers $(O_i, O)$ of the image planes, then the desired joint angle vector $Q_{Oi}$, which is necessary to bring $V_i$ to $V_{Oi}$, can be computed as follows,

$P_i = (V_{Oi}, Q_{Oi})$, (3)

$Q_{Oi} = Q_i [V_i, P_i]$, (4)

where and are the inversion functions which can be used to control the fixation on $p_i$, respectively. Because $P_i$ has invariance, $Q_{Oi}$ can be computed by combining Eqs.(2) and (3) with Eq.(4),

$Q_{Oi} = -1 [V_{Oi}, (V_i, P_i)]$, (5)

therefore, $Q_{Oi}$ is used to control the active vision system to fixate $p_i$.

4.2 Grasping Control

Figure 3 shows the configuration parameters of the active stereo vision system. Let $d_s, l$ be the distance, height of the CCD cameras, diameter of the sphere coordinate system, $i = \alpha, \beta, \gamma$ be the configuration angles of the active stereo vision system, the spatial coordinates of $p_i$ in are $c = [x_{zi}, y_{zi}, z_{zi}]$.
respectively. When the active stereo vision system fixes \( p_i \), \( c_p \), can be computed by the triangular geometry relationships in Fig. 3.

\[
\begin{align*}
X_{ci} &= X_{ci} = \cos(\theta_i) \sin(\phi), \\
Y_{ci} &= Y_{ci} = \cos(\phi), \\
Z_{ci} &= Z_{ci} = \sin(\theta_i).
\end{align*}
\]

Where \( \mathbf{W}(t) \) is the original coordinates of \( \nu \) in \( \mathcal{O} \), \( \mathbf{J}[\nu(t)] \in \mathbb{R}^{6 \times 1} \) is a Jacobian matrix of the end-effector. Therefore, we have

\[
\dot{\nu}(t) = \mathbf{J}[\nu(t)]^T \dot{\mathbf{w}}(t).
\]

When the sampling period of the robot joint control system \( T \) is very minute, it is suitable that using \( \nu(k+1) = \nu(k) + \frac{\dot{\nu}(k)}{T} \) to replace \( \nu(k) \) at time \( t = kT \), therefore, Eq. (11) is discreted by

\[
\nu(k+1) = \nu(k) + \frac{\dot{\nu}(k)}{T}.
\]

By homogeneous transformation relationship, \( \mathbf{w}_p \) can be specified by

\[
\begin{align*}
\mathbf{w}_p &= \mathbf{h}_c^T \mathbf{c}_p, \\
\mathbf{p}_b &= \mathbf{h}_b^T \mathbf{p}_c, \\
\mathbf{p}_c &= \mathbf{h}_c^T \mathbf{p}_b,
\end{align*}
\]

Where \( \mathbf{h}_c, \mathbf{h}_b, \mathbf{h}_6, \mathbf{h}_0, \mathbf{h}_b, \mathbf{h}_0 \) are the homogeneous matrices from \( c \) to \( b \), \( \mathbf{h}_c \) to \( c \), \( \mathbf{h}_b \) to \( b \), respectively.

According to the robotic forward kinematics \( \mathbf{t} \), \( \mathbf{r} \), is a reference joint angle vector of the end-effector. we obtain

\[
\begin{align*}
\dot{W}(t) &= \mathbf{J}[
u(t)], \\
\mathbf{J}[
u(t)] &= \frac{\partial \mathbf{t}}{\partial \nu}, \\
\dot{\mathbf{w}}(t) &= \mathbf{J}[
u(t)]^T \dot{\nu}(t).
\end{align*}
\]

5. Visual Robot Learning Control System

5.1 Visual Learning Control System

In order to obtain the nonlinear many-to-one mapping function, ART_NN are combined with FF_NN to learn defined in Eq. (2). The architecture of ART_NN, FF_NN and the vision-based robot control system based on ART_NN and FF_NN are showed in Figs. 5 and 6, respectively.
the self-adaptive resonance learning algorithms for ART_NN and the error back propagation learning algorithms for FF_NN are omitted. In Fig.6, the ART_NN require two types of inputs $V_i$ and $Q_i$, where $V_i$ corresponds to the image coordinates of $p_i$ on the image planes of the CCD cameras, and $Q_i$ is the joint angle coordinates corresponding to the CCD cameras tracking $p_i$. The ART_NN clusters $V_i$ into classes within the category layer. The class number in each category layer depends on a vigilant parameter which is a real number between zero and one.

![Fig.6 A Robotic Learning Control System](image)

In Fig.6, $K_1$ and $K_2 \in R^{6 \times 6}$ are the coefficient matrices which were specified empirically, $E_{le}$ and $E_{re} \in R^{6 \times 6}$ are the differences between the two learning trials, respectively, and the PID controller is used to obtain the joint servoing control with high accuracy.

6. Simulations
To evaluate the validity of the active stereo vision-based robotic tracking, fixating and grasping in the unknown environment, the simulation are carried out using the models of the active stereo vision system installed in the end-effector.

For controlling the robot to track, fixate and grasp the object, first, ART_NN1 and ART_NN2 learn the many-to-one functional mapping relationships by generating 10000 random pairs of $V_i$ and $Q_i$ signals corresponding to $p_i$. The ART_NN1 created 500 classes for the inputs from the right camera, and the ART_NN2 also created 500 classes for the inputs from the left camera. Second, the spatial coordinates of $p_i$ and the desired joint angles $Q_{wi}$ are computed and used in the tracking, fixating and grasping control loop.

The simulation results denotes that the errors for all of the three components of $O_p$, converged to within 2% of its dynamic range. These results show that the learning of ART_NN and FF_NN is fast and convergent, and the end-effector can also arrive at the position of the object and grasp it.

7. Conclusions
The following conclusions have been obtained from the above experiments:
1. There exist many-to-one mapping relationships between the joint angles of the active stereo vision system and the spatial representations of the object in the workspace frame.
2. ART_NN and FF_NN can learn the mapping relationships in an invariant manner to the changing joint angles, the vision and joint angle signals of the active vision system corresponding to the object correspond to the same spatial representation of the object.
3. The present approach was evaluated by simulation using the models of the 11 DOF active stereo vision system, the simulation confirms that the present approach has high robustness and stability.

8. Acknowledgement
The research in this paper is founded by the initial foundation of China Education Ministry for the Scholars Returned China from abroad (Project No.: 2002247), the authors are greatly grateful to the foundations.

References

