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-effecter 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-effecter 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 $_{o}$. According to the visual feed back information, the robot can track, fixate and grasp autonomously the object.

Authorized licensed use limited to: IEEE Xplore. Downloaded on March 13, 2009 at 00:19 from IEEE Xplore. Restrictions apply.

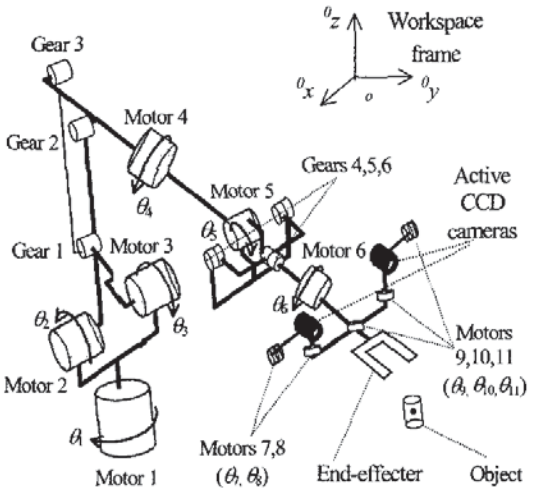


Fig.1 A robot system with active vision

3. Many-to-One Mapping Relationships

Figure 2 shows the projective relationships between the active stereo vision system and the object in $_o$. 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_l, o_r) of the left and right image planes of the two CCD cameras, respectively, the active stereo vision system is known as fixation on p_i.

Let $Q_i=[\ _{i7},\ _{i8},\ldots,\ _{i11}]^T$, ${}^OP_i=[x_{oi}, y_{oi}, z_{oi}]^T$ and $V_i=[{}^l u_i, {}^l v_i, {}^r u_i, {}^r{}_i]^T$ be a joint angle vector of the active stereo vision system tracking p_i, a spatial representation vector of p_i in $_o$ and an image coordinate vector of p_i on the left and 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 ${}^OP_2$ of p_2 identified by the active stereo vision system tracking on p_2 should be different from Q_1 and ${}^OP_1$. If another joint angle vector Q_3 is obtained by tracking p_3, p_1 and p_2 are still visible, ${}^OP_1$ and ${}^OP_2$ 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 ${}^OP_i$, which means that ${}^OP_i$ is invariant to the changing Q_i and V_i.

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

$$V_i = (Q_i, {}^OP_i), \quad (i=1,2,3), \qquad (1)$$

where 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, ${}^OP_i$ is specified as

$${}^OP_i = (V_i, Q_i), \quad (i=1,2,3), \qquad (2)$$

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

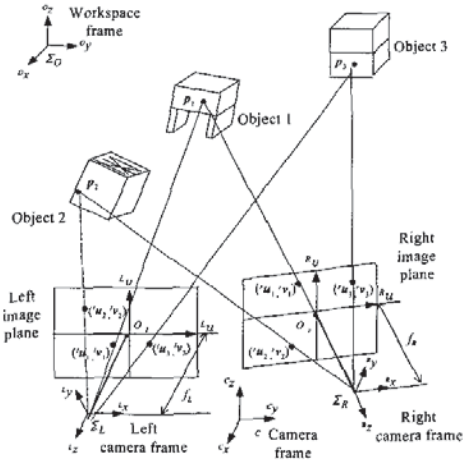


Fig.2 Projection and fixation relationships

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 ${}^OP_i$ and Q_i, we have from Eq.(2), ${}^OP_i = (V_i, Q_i)$, where the active stereo vision system tracks p_i. When the active stereo vision system fixates p_i and the image coordinate vector V_{Oi} of p_i corresponds to the centers (o_l, o_r) 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,

$${}^OP_i = (V_{Oi}, Q_{Oi}), \qquad (3)$$

or

$$Q_{Oi} = {}^{-1}(V_{Oi}, {}^OP_i), \qquad (4)$$

where and ${}^{-1}$ are invertible functions which can be used to control the fixation on p_i, respectively. Because ${}^OP_i$ has invariance, Q_{Oi} can be computed by combining Eqs.(2) and (3) with Eq.(4),

$$Q_{Oi} = {}^{-1}[V_{Oi}, (V_i, Q_i)], \qquad (5)$$

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, $_{i\alpha}$, $_{i\beta}$, $_{i\gamma}$ be the configuration angles of the active stereo vision system, the spatial coordinates of p_i in $_c$ be ${}^cP_i=[x_{ci}, y_{ci}, z_{ci}]^T$,

Authorized licensed use limited to: IEEE Xplore. Downloaded on March 13, 2009 at 00:19 from IEEE Xplore. Restrictions apply.

respectively. When the active stereo vision system fixates p_i, ${}^{C}P_i$ can be computed by the triangular geometry relationships[1] in Fig.3

$$x_{ci} = l \cos(\ _{i\gamma}) \sin(\ _{i\beta}), \quad (6.a)$$

$$y_{ci} = l \cos(\ _{i\gamma}) \cos(\ _{i\beta}), \quad (6.b)$$

$$z_{ci} = l \sin(\ _{i\gamma}). \quad (6.c)$$

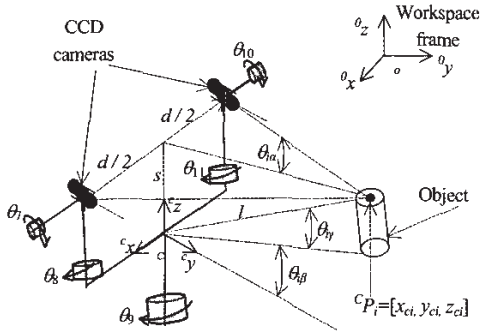


Fig.3 Configuration of the vision system

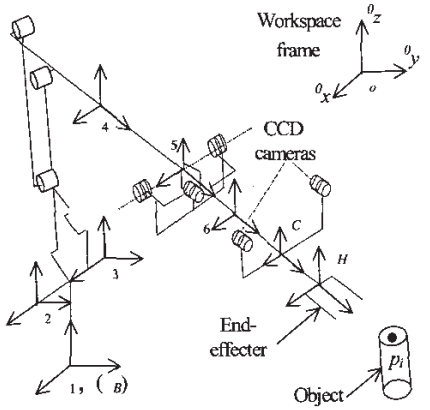


Fig.4 Frames of the active vision system

Figure 4 gives the joint coordinate frames of the robot joint system. In Fig.4, let $_j (j=1,2,\ldots,6)$, $_B$, Σ_H, Σ_C be a coordinate frame of ith robotic joint, base frame, coordinate frame of the end-effecter, camera frame, ${}^{B}P_i = [x_{bi}, y_{bi}, z_{bi}]^T$, ${}^{H}P_i = [x_{hi}, y_{hi}, z_{hi}]^T$ be an Euclidian coordinate vector of p_i in $_B$, $_H$ respectively.

By homogeneous transformation relationship, ${}^{O}p_i$ can be specified by

$${}^{H}P_i = {}^{H}H_C \cdot {}^{C}P_i, \quad (7.a)$$

$${}^{B}P_i = {}^{B}H_6 \cdot {}^{6}H_H \cdot {}^{H}P_i, \quad (7.b)$$

$${}^{O}P_i = {}^{O}H_B \cdot {}^{B}P_i. \quad (7.c)$$

where ${}^{H}H_C$, ${}^{6}H_H$, ${}^{B}H_6$, ${}^{O}H_B$ are the homogeneous matrixes from $_C$ to $_H$, $_H$ to $_6$, Σ_6 to $_B$, $_B$ to $_O$, respectively.

According to the robotic forward kinematics $[\ _r(t)] \in R^{6\times 1}$, $_r = [\ _1, \ldots, \ _6]^T$ is a reference joint angle vector of the end-effecter. we obtain

$${}^{O}W(t) = [\ _r(t)], \quad (8)$$

$$J[\ _r(t)] = \partial [\ _r(t)] / \partial\ _r(t), \quad (9)$$

$${}^{O}\dot{W}(t) = J[\ _r(t)]\ \dot{}_r(t), \quad (10)$$

Where ${}^{O}W(t)$ is the original coordinates of $_H$ in $_O$, $J[\ _r(t)] \in R^{6\times 1}$ is a Jacobian matrix of the end-effecter. Therefore, we have

$$\dot{}_r(t) = J^{-1}[\ _r(t)] \cdot {}^{O}\dot{W}(t). \quad (11)$$

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

$$[\ _r(k+1) - \ _r(k)]/T = J^{-1}[\ _r(k)] \cdot {}^{O}\dot{W}(k), \quad (12.a)$$

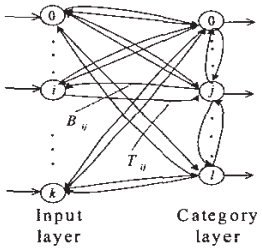
or $$\ _r(k+1) = \ _r(k) + TJ^{-1}[\ _r(k)] \cdot {}^{O}\dot{W}(k), \quad (12.b)$$

where ${}^{O}\dot{W}(k) = [{}^{O}W(k) - {}^{O}W(k-1)]/T$, $_r(k+1)$ can be used to control the robot joint angles. When ${}^{O}W(t) = {}^{O}P_i$, the end-effecter can grasp the object.

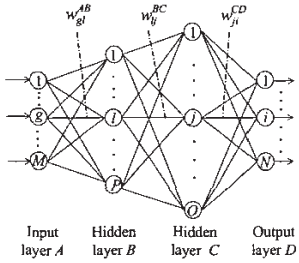
5. Visual Robot Learning Control System

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.



(a) Architecture of ART_NN



(b) Architecture of FF_NN

Fig.5 Architectures of ART_NN and FF_NN

5.2 Learning of ART_NN and FF_NN

In Fig.5, T_{ij}, B_{ij}, w_{gl}^{AB}, w_{lj}^{BC} and w_{ji}^{CD} are weights,

Authorized licensed use limited to: IEEE Xplore. Downloaded on March 13, 2009 at 00:19 from IEEE Xplore. Restrictions apply.

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.

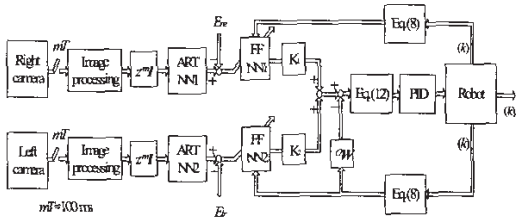


Fig.6 A Robotic Learning Control System

In Fig.6, $\boldsymbol{K}_1$ and $\boldsymbol{K}_2 \in R^{6\times 6}$ are the coefficient matrixes which were specified empirically, $\boldsymbol{E}_{le}$ and $\boldsymbol{E}_{re} \in R^{6\times 6}$ 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-effecter.

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_{oi} 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 ${}^{O}P_i$ 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-effecter 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

(1) Bernardino A., Santos-Victor, Binocular Tracking: Integrating Perception and Control, *IEEE Transactions on Robotics and Automation*, (1999-12), Vol. 15, No. 6, pp. 1080 -1093.

(2) Sumi, Kawai, Yoshimi, 3-D Object Recognition in Cluttered Environments by Segment-Based Stereo Vision, *International Journal of Computer Vision*, (2002-1), Vol. 46, No. 1, pp. 5-23.

(3) Srinivasa N., Rajeev S., SOIM: A Self-Organizing Invertible Map with Applications in Active Vision, *IEEE Transactions on Neural Networks*, (1997-5), Vol. 8, No. 3, pp. 758-773.

(4) Barnes N.M., Liu Z.Q., Vision guided circumnavigating autonomous robots, *International Journal of Pattern Recognition and Artificial Intelligence*, (2000), Vol. 14, No. 6, pp. 689-714.

Authorized licensed use limited to: IEEE Xplore. Downloaded on March 13, 2009 at 00:19 from IEEE Xplore. Restrictions apply.