MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

Hand-Eye Coordination Using Active Stereo Camera

WeiYun Yau Han Wang Dinesh P. Mital
School of Electrical & Electronic Engineering
Nanyang Technological University, Singapore

Abstract

This paper describes an approach to control the robot arm using active stereo camera system. The proposed hand-eye system is able to achieve high accuracy without drastically reducing the workspace size. A new qualitative approach to control the robot arm is developed. It computes the relative depth between two points from a pair of stereo image. By incorporating this attribute into the image space, a pseudo three-dimensional (3D) image space is obtained. Subsequently, the pseudo image space is used to compute the required transformation from the image space to the robot space. Such an approach does not require the recovery of the intrinsic and extrinsic parameters of the stereo vision system or the 3D coordinates of the target object. Therefore, it is robust to changes in the parameters of the vision system and thus allows the integration of active vision system. A method to cater for focal length changes for achieving variable resolution is also described. Experiments are conducted to verify the accuracy and performance of the proposed method.

1 Introduction

One of the most important tasks that the human visual system engages in is hand-eye coordination. Hand-eye coordination has been and active area of research recently. In general, the hand-eye coordination system can be divided into (a) eye-in-hand and (b) eye-to-hand configuration. In the former, the vision system is mounted on the robot arm while in the latter, the vision system is separated from the robot arm. In this paper, only the eye-to-hand configuration will be described. For such configuration, almost all researchers employ passive cameras to control the robot arm. Most algorithms require the vision system to be calibrated in order to recover the 3D world with respect to the robot frame. However, such approach is not practical for use with the active vision system as it involves re-calibration of the vision system whenever any parameter of the vision system is changed. Methods that does not require recovery of 3D structure are proposed in [2, 3, 1] but they do not address the issues of active vision control. For passive cameras, the accuracy and workspace size of the hand-eye system are limited by the stereo vision system as usually the robot has better accuracy and working range. Such bottleneck can be overcomed by the use of active vision. The main goal of this paper is to show how an active stereo camera can be used to control the robotic arm. For passive camera system, increasing the accuracy inevitably decreases the workspace size. The approach proposed in this paper is able to achieve high accuracy and large workspace size. During the search stage, wider view angle is used and at the manipulation stage, the vision system can fixate at the object to be manipulated and increase its resolution without losing sight of the object.

2 Qualitative Approach

Hand-eye coordination does not require the precise quantitative recovery of the 3D world coordinates to carry out most of its tasks successfully. Qualitative and relative information will suffice, even for accurate positioning. The advantage of such an approach is that it is robust to changes in the visual parameters, thus allowing the integration of active vision system. The most crucial problem of hand-eye coordination is depth recovery. Instead of using absolute depth, the use of relative depth obtainable from a stereo pair images is proposed. Consider a general stereo camera platform with configuration as shown in Fig. 1. A general stereo camera platform is where the vergence angle[1] is between 0° and 180° non-inclusive. Consider a small cyclo-torsion angle, ϕ, a small tilt angle difference, β, and a small vertical offset, dy, between the right and the left camera.

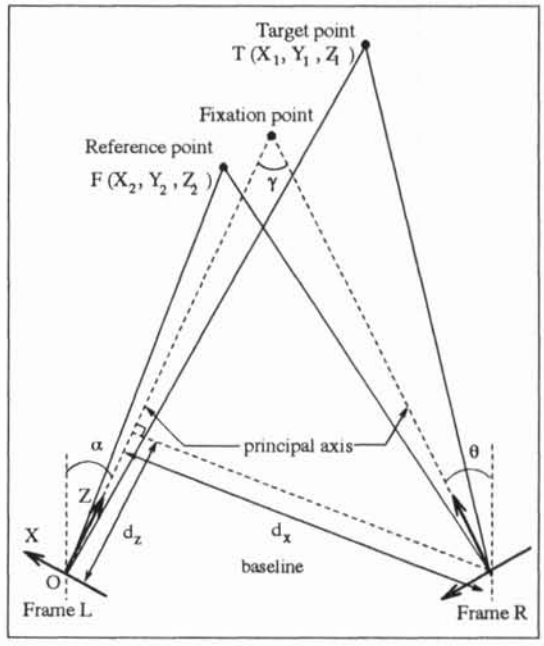

Figure 1: Geometry of the general stereo camera configuration.

[1] angle between both principle axes on the plane defined by the two optical centers and the fixation point.

Let the pan angle of the left camera be α and that of the right camera be θ. Consider two world points, the reference point F(X_f, Y_f, Z_f) and target point, T(X_t, Y_t, Z_t). Their image coordinates are (u_f, v_f) and (u_t, v_t) respectively. Define **relative stereo disparity**, rsd, as the difference in the disparity of the reference and target point in the reference frame with the disparity of the reference and target point in the other stereo frame scaled by their respective finite vertical disparities.

$$rsd = (u_{lt} - u_{lf}) - (\frac{v_{lt}}{v_{rt}} u_{rt} - \frac{v_{lf}}{v_{rf}} u_{rf}) \quad (1)$$

where subscript l and r denotes the left and right frame respectively. By expanding equation (1) using perspective projection and linearizing it, retaining only the first order term in the process, it can be shown that the rsd has the following relation (the derivation can be found [4]).

$$rsd = \frac{f\alpha_u}{Z_f Z_t}(d_x \frac{\cos\gamma}{\cos\beta} - d_z \sin\gamma)(Z_t - Z_f) \quad (2)$$

where α_u is the inverse of the horizontal pixel size (in units of length, e. g. meters). Equation (2) shows a strikingly simple form of the relative depth. The relation of rsd with the relative depth is monotonic. The absolute value tells the depth-wise relation while the sign tells the relative arrangement between the two points. Note that using Equation (1), the accuracy of the positioning achievable is independent of calibration. This is because from perspective projection, two points in the general stereo camera configuration will coincide in the world space if and only if there exists no relative disparity in the images of both cameras simultaneously.

A pseudo 3D image space can be obtained by incorporating the rsd into the 2D image space as the third dimension. This is possible because the rsd measures the relative depth, a dimension which is not coplanar with the image plane that forms the other two dimensions. The pseudo image space has the same dimension as the world space and a linear relation between them can be obtained. By choosing the world space to be the robot space, the required hand-eye transformation can be computed. Define the horizontal error and vertical error as the difference in the horizontal and vertical coordinates between the target and reference points. The horizontal error is given by $(u_t - u_f)$ while the vertical error is given by $(v_t - v_f)$. Thus, the pseudo image error vector is given by the vector $[u_t - u_f, v_t - v_f, rsd]^T$. Assume that the two points have small relative depth error, then the horizontal and vertical errors in the image space projected to the camera coordinate frame can be obtained by using the affine projection. The transformation between the pseudo image space to the robot space is given by the following equation.

$$\begin{bmatrix} u_t - u_f \\ v_t - v_f \\ rsd \end{bmatrix} = \mathbf{D}\,\mathbf{R} \begin{bmatrix} X_t - X_f \\ Y_t - Y_f \\ Z_t - Z_f \end{bmatrix} \quad (3)$$

where

$$\mathbf{D} = \begin{bmatrix} \alpha_u f/Z_t & 0 & 0 \\ 0 & \alpha_v f/Z_t & 0 \\ 0 & 0 & \alpha_u f n/Z_f Z_t \end{bmatrix}$$

$\mathbf{R}(r_{ii})$; $i = 1, 2, 3$ is the rotation matrix from the camera coordinate frame to the robot coordinate frame, $n = d_x \cos\gamma/\cos\beta - d_z \sin\gamma$ and α_v is the inverse vertical pixel size.

The above equation relates the pseudo 3D image space to the 3D robot space. This linear model provides a qualitative value which indicates nearness when the two points concerned are not close to each other (not localized). This information can be used to navigate the point F towards the point T. As both points are localized, the values obtained can be considered quantitatively. This will allow F to be guided accurately to reach T. By implementing equation (3) to solve for the hand-eye coordination, camera calibration to recover the intrinsic and extrinsic parameters of the stereo camera is not needed. The required image-to-robot transformation matrix can be easily computed online by letting the end-effector perform three orthogonal movements. It is a square matrix of dimension three, thus the computation required is inexpensive. Further incorporating visual feedback to update the transformation matrix regularly gives the hand-eye system robustness to changes in the stereo camera configuration [4]. It allows the active vision system to fixate at any location in the robot's workspace, maximizing the robot's capability. Therefore, active vision system can be incorporated without having to re-calibrate the hand-eye system or requiring extensive computations to recover the required parameters.

3 Focal Length Changes

One of the factor that affects the accuracy of the hand-eye system is the focal length. By using motorized zoom and focus lenses in an active vision setup allows the resolution of the hand-eye system to be dynamically controlled. This has the advantage in that during the search stage, a smaller focal length (wide angle) is used so that the field of view of the stereo vision system is sufficiently large for the target object to be promptly and easily located. However, the image resolution may not be sufficient for the end-effector to perform the required task. As the reference point is approaching the target point, the focal length can be increased gradually. This reduces the field of view but increases the resolution of the stereo camera system.

Decoupling the focal length term from equation (3) and simplifying gives the following linear equation.

$$\mathbf{u} = f\mathbf{M}\mathbf{w} \quad (4)$$

where

$$\begin{aligned} \mathbf{u} &= (u_t - u_f, v_t - v_f, rsd)^T \\ \mathbf{w} &= (X_t - X_f, Y_t - Y_f, Z_t - Z_f)^T \\ \mathbf{M} &= \begin{bmatrix} \alpha_u r_{11}/Z_t & \alpha_u r_{12}/Z_t & \alpha_u r_{13}/Z_t \\ \alpha_v r_{21}/Z_t & \alpha_v r_{22}/Z_t & \alpha_v r_{23}/Z_t \\ n r_{31}/Z_f Z_t & n r_{32}/Z_f Z_t & n r_{33}/Z_f Z_t \end{bmatrix} \end{aligned}$$

When the focal length is changed to a new value, f', the pseudo image error vector will be changed too. Performing some simple algebraic manipulation gives the following equation.

$$\mathbf{u}' = kf\mathbf{M}\mathbf{w} \quad (5)$$

where $k = f'/f$. From the equation (5), only the zoom factor, k, need to be computed whenever the focal length is changed. The zoom factor can be known from the lens modeling or by calculating the ratio of the image size before and

after the change in the focal length. Note that the actual focal length value need not be known and hence calibration to recover the focal length is not necessary. Furthermore, error in computing the zoom factor is much smaller compared to the actual recovery of the focal length. Another point worth mentioning is that according to equation (1), the rsd only depends on the focal length of the reference camera. Small mismatch in the focal length of the two lenses will be taken care of by the ratio of the vertical disparities.

4 Experiments

To test the accuracy of the active hand-eye system, we let the end-effector of the robot hold a 3.5 inch floppy disk, called the reference disk. Another similar floppy disk, the target disk, is arbitrarily placed in the workspace of the robot. The task of the hand-eye system is to align the bottom-left corner (reference corner) of the reference disk to the top-right corner (target corner) of the target disk [1]. The corners are tracked and their coordinates are fed back to the main controller to control the robot arm and update the transformation matrix. As the target and reference corners are close to each other, the visual feedback is disabled. The robot arm then performs a one shot movement to reach the target corner.

Two set of tests were conducted. In the first set, the stereo cameras were stationary and the focal length was preset to 25mm. The robot was then activated to align the reference corner to the target corner. Upon completion, any position error was recorded by manually offsetting the error using a teach pendant. The alignment task was repeated for increasing focal lengths of 35mm and 45mm. The initial focal length was still set to the preset value of 25mm, but as the end-effector moved towards the target disk, the focal length was increased to the required value. The test was then repeated for the second set where the pan-tilt units were activated to fixate the stereo cameras at the target corner. The fixation process were activated only after the end-effector has moved towards the target disk. Once the two sets were completed, the position of the target disk was changed and the whole process was repeated. A total of 50 readings were taken for each focal length and the statistics of the results obtained are provided. The largest positive and negative errors detected are presented in Fig. 2 and Fig. 3 respectively while the mean error and the standard deviation are shown in Fig. 4 and Fig. 5. Note that positive value of the error indicates overshoot.

4.1 Discussion

The results obtained in the accuracy test for the case of static camera and fixating camera system as shown in Figures 2, 3, 4 and 5 reveal that fixation has negligible effect on the performance of the hand-eye coordination. Both static and fixating system show improvement in the accuracy as the focal length is increased. The gain in accuracy from the increase in the focal length far exceeds the error due to fixation, if any. Therefore, the advantages of using the active camera system become clear. It increases the workspace of the hand-eye system as well as its accuracy.

For an ideal system setup, there should not be any position error in the alignment of the corners as proven in the previous section. Any error in the results obtained must be mainly due to the physical limitation of the system. The maximum error arising from the physical system used is estimated using a baseline of 940mm and the maximum depth of the target point from the baseline at 2050mm. From the specification of the camera and assuming an error of one pixel, the expected maximum vertical, horizontal and depth positioning errors for all the focal lengths used are shown in Fig. 2 and Fig. 3 for comparison. Analyzing these results, it can be concluded that the depth error obtained is within the expected limit since the corners are tracked up to sub-pixel accuracy. However, the horizontal and vertical depth exceeds the expected limit. This is because the actual corners of the floppy disks are rounded. During manual alignment, the corners are aligned such that the two rounded corners touch each other to reduce inconsistency. This causes some offset as the corners detected are extrapolated. However, such offset has little effect on the depth accuracy as the rsd depends on the relative separation and not on the absolute position of the corner. As long as the corner can be consistently localized, the depth accuracy will be good. Furthermore, to avoid the reference corner from occluding the target corner, vertical offset is included before the final alignment. Inaccuracies may arise in removing the vertical offset during the final alignment, which explains why the vertical error is usually larger than the horizontal error though the calculated values show the opposite. We would like to emphasize that in the final alignment, the visual feedback is disabled. The conformity of the obtained results with the expected accuracy computed suggest that the use of the pseudo image space and the resulting transformation matrix is acceptable for solving the hand-eye coordination problem.

5 Conclusions

In this paper, we have presented an approach to control the robot arm using the active vision system to achieve the active hand-eye coordination system. The advantage of such a system is that it increases the flexibility and the workspace size of the hand-eye system without compromising the achievable accuracy. The use of fixation allows focal length to be increased to achieve good accuracy, sufficient for the required manipulation task. The proposed method does not require the recovery of the intrinsic and extrinsic parameters of the stereo vision system or the 3D coordinates of the target object. Furthermore, the algorithm is simple and fast, making the algorithm suitable for real-time visual feedback implementation. Although there are still many unanswered research issues, we believe this work will be an impetus towards the successful development of a well coordinated active head-eye-hand system which seems effortless in all animals especially the human beings.

References

[1] G.D. Hager, W.C. Chang, and A.S. Morse. Robot hand-eye coordination based on stereo vision. *IEEE Control Systems*, pages 30–9, February 1995.

[2] N. Hollinghurts and R. Cipolla. Uncalibrated stereo hand-eye coordination. *Image and Vision Computing*, 12(3):187–92, 1994.

[3] K. Hosoda and M. Asada. Versatile visual servoing without knowledge of true jacobian. In *Proceedings International Conference on Intelligent Robots and Systems*, volume 1, pages 186–93, 1994.

[4] W.Y. Yau and H. Wang. Robust hand-eye coordination. *Advanced Robotics*, Feb 1996. submitted for publication.

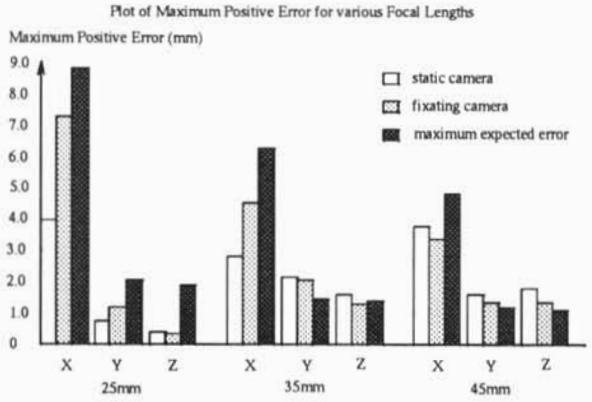

Figure 2: Maximum positive error for various focal lengths.

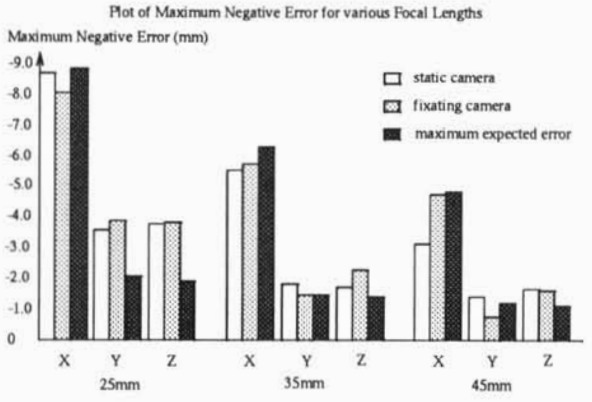

Figure 3: Maximum negative error for various focal lengths.

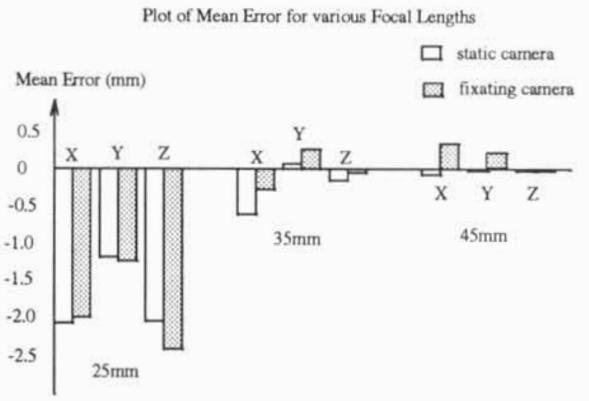

Figure 4: Mean error obtained for various focal lengths.

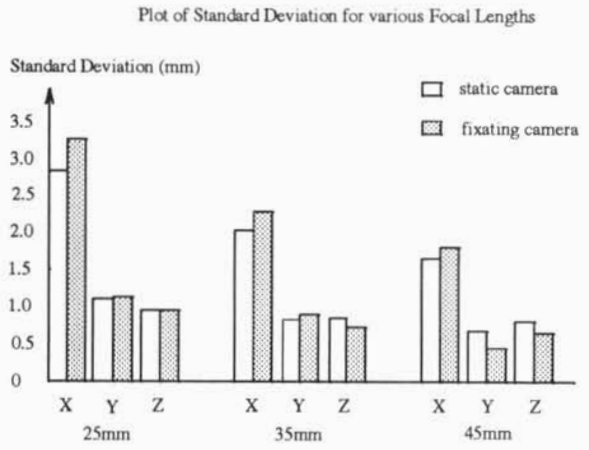

Figure 5: Standard deviation of error obtained for various focal lengths.