In *Proc. of the International Federation of Automatic Control (IFAC) 1993 World Congress*, Sydney, Australia, July 18-23, 1993.

DYNAMIC SENSOR PLACEMENT USING CONTROLLED ACTIVE VISION

BRAD NELSON*, N. P. PAPANIKOLOPOULOS and PRADEEP K. KHOSLA***

**The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA*

***Department of Computer Science, University of Minnesota, 200 Union St. SE, Minneapolis, MN 55455, USA*

Abstract. The working region of a camera providing visual feedback to guide manipulation tasks can be greatly extended if the camera can move in real-time. Proper camera motion depends on several different criteria, such as object motion, maximum manipulator tracking speed, manipulator configuration, camera depth-of-field, field-of-view, spatial resolution, occlusion avoidance, etc. In this paper, the controlled active vision framework is extended to include dynamically determined sensor placement criteria. The criteria considered include focus, spatial resolution, and manipulator configuration, as well as object motion. Experimental results are presented.

Key Words. assembly; automation; manipulation; real-time control; robots; visual tracking; visual servoing

1. INTRODUCTION

Visual feedback can complement force feedback in imprecisely calibrated robotic workcells so that precise manipulation tasks can be successfully performed. A useful technique for increasing the working region of a camera providing this visual feedback is to mount the camera on the end-effector of a manipulator. The manipulator holding the camera becomes an eye-in-hand system, and control issues associated with these types of visual servoing systems must be addressed. In addition to eye-in-hand control issues, the placement of the camera in the work cell must be considered. Assuming a camera with a fixed focal length lens is used, the eye-in-hand system can be used to move the camera closer to the assembly task being performed in order to increase the spatial resolution of the sensor such that the static visual servoing algorithm can servo on the object to a sufficient accuracy. Another benefit of moving the camera is to change the viewing direction with which the camera observes the object so that the spatial resolution in directions formerly along the optical axis is increased. However, the object being tracked must remain in focus and within the field-of-view of the camera. When servoing a camera mounted at a manipulator's end-effector, it is also important that the manipulator holding the camera maintains a "good" configuration far enough from kinematic singularities so that manipulator cartesian control algorithms are properly conditioned.

In the past, camera placement has been determined by considering such criteria as occlusions, field-of-view, depth-of-field, and/or camera spatial resolution (Cowan and Bergman, 1989; Tarabanis *et al*., 1991; Yi *et al*., 1990). In none of these cases, however, is the camera actually servoed based on visual data. For dynamically changing manipulation tasks, the camera must move in real-time, so the placement of the camera must be determined quickly, therefore, visual tracking algorithms can be effectively applied. The configuration of the manipulator must also be taken into account in the control strategy, unless the working region of the workcell can be severely constrained.

The controlled active vision paradigm (Papanikolopoulos *et al*., 1991) is a useful framework for incorporating sensor placement criteria like those previously mentioned into an eye-in-hand robotic system. In this paper, the controlled active vision framework will first be used to derive a system model and controller for an eye-in-hand system. Some dynamically determined sensor placement criteria will be presented, and it will then be shown how the control objective function can be augmented in order to introduce these various sensor placement criteria into the visual tracking control law. A brief description of the experimental system and presentation of experimental results complete the paper.

2. MODELING AND CONTROL

The controlled active vision framework is used to model the eye-in-hand system (Papanikolopoulos *et al*. 1991). In formulating the visual tracking model, a full 3D tracking model is created so that less complex tracking strategies can be more easily compared.

2.1. Modeling

To model the 3-D visual tracking problem, a pinhole model for the camera with a frame {C} placed at the focal point of the lens is used, as shown in Fig. 1. A feature on an object with coordinates (X_o, Y_o, Z_o) in the camera frame projects onto the camera's image plane at (x,y). The manipulator holding the camera provides camera motion by moving the camera frame along its X,Y,Z axes. The eye-in-hand visual tracking system can be written in state-space form as

$$\boldsymbol{x_F}(k+1) = \boldsymbol{A_F}\boldsymbol{x_F}(k) + \boldsymbol{B_F}(k)\boldsymbol{u}(k) + \boldsymbol{E_F}\boldsymbol{d_F}(k) \qquad (1)$$

where $\boldsymbol{A_F}=\mathbf{I}_2$, $\boldsymbol{E_F}=T\mathbf{I}_2$, $\boldsymbol{x_F}(k)\in R^2$, $\boldsymbol{d_F}(k)\in R^2$, and $\boldsymbol{u}(k)\in R^6$. The matrix $\boldsymbol{B_F}(k)\in R^{2\text{x}6}$ is

$$\boldsymbol{B_F}(k) = T\begin{bmatrix} -\frac{f}{Z_o(k)s_x} & 0 & \frac{x(k)}{Z_o(k)} & \frac{x(k)y(k)s_y}{f} & -\left(\frac{f}{s_x}+\frac{x^2(k)s_x}{f}\right) & \frac{y(k)s_y}{s_x} \\ 0 & -\frac{f}{Z_o(k)s_y} & \frac{y(k)}{Z_o(k)} & \left(\frac{f}{s_y}+\frac{y^2(k)s_y}{f}\right) & -\frac{x(k)y(k)s_x}{f} & -\frac{x(k)s_x}{s_y} \end{bmatrix} \qquad (2)$$

The vector $\boldsymbol{x_F}(k)=[x(k) \ \ y(k)]^{\mathrm{T}}$ is the state vector, $\boldsymbol{u}(k)=[\dot{x}_c \ \ \dot{y}_c \ \ \dot{z}_c \ \ \omega_{xc}\omega_{yc}\omega_{zc}]^{\mathrm{T}}$ is the vector representing possible control inputs, and $\boldsymbol{d_F}(k)$ is the exogenous deterministic disturbances vector due to the feature's optical flow induced by object motion. The state vector $\boldsymbol{x_F}(k)$ is computed using the SSD algorithm to be described in Section 2.3. In (2), f is the focal length of the lens, s_x and s_y are the horizontal and vertical dimensions of the pixels on the CCD array, and T is the sampling period between images. In order to simplify notation without any loss of generality, let $k=kT$. In addition, it is assumed that $Z_o >> f$. This assumption holds because the focal length of our camera is 20mm, while Z_o is larger than 500mm.

Depending on the constraints placed on target motion and the objective of the visual tracking system, more than one feature may be required in order to achieve the system's goals. For example, for full 3D tracking in which it is desired to maintain a constant six degree of freedom transformation between the camera and the target, at least three features are required. To track an object constrained to move with motion in three dimensions, such as planar motion with rotations or 3D translational motion, at least two features are needed. A generalized state equation for a variable number of features can be written as

$$\boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}(k)\boldsymbol{u}(k) + \boldsymbol{E}\boldsymbol{d}(k) \qquad (3)$$

where M is the number of features required, $\boldsymbol{A}=\mathbf{I}_{2M}$, $\boldsymbol{E}=T\mathbf{I}_{2M}$, $\boldsymbol{x}(k)\in R^{2M}$, $\boldsymbol{d}(k)\in R^{2M}$, and $\boldsymbol{u}(k)\in R^i$ ($i\in\{1,2,3,4,5,6\}$, the number of axes along which tracking occurs). The matrix $\boldsymbol{B}(k)\in R^{2M\text{x}i}$ is

$$\boldsymbol{B}(k) = \begin{bmatrix} \boldsymbol{B_F}^{(1)}(k) \\ \ldots \\ \boldsymbol{B_F}^{(M)}(k) \end{bmatrix} \qquad (4)$$

The superscript (j) denotes each of the feature points ($j\in\{1,2,...,M\}$). Thus, the size of $\boldsymbol{B}$ is dependent on the number of non-zero cartesian control inputs and the number of features required, which the system designer determines based on task requirements. The vector $\boldsymbol{x}(k)=[x^{(1)}(k) \ y^{(1)}(k) \ldots \ x^{(M)}(k) \ y^{(M)}(k)]^{\mathrm{T}}$ is the new state vector, and $\boldsymbol{d}(k)$ is the new exogenous deterministic disturbances vector.

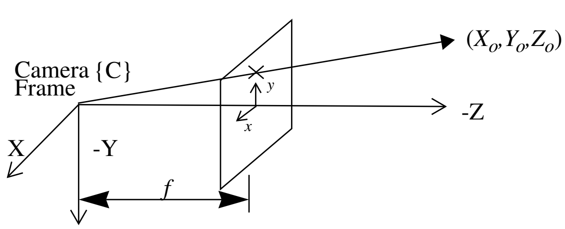

Fig. 1. The pinhole camera model.

2.2. Control

The control objective of an eye-in-hand visual tracking system is to control camera motion in order to place the image plane coordinates of features on the target at some desired position, despite object motion. The desired image plane coordinates could be changing with time, or they could simply be the original coordinates at which the features appear when tracking begins. The control strategy used to achieve the control objective is based on the minimization of an objective function at each time instant. The objective function places a cost on differences in feature positions from desired positions, as well as a cost on providing control input, and is of the form

$$F(k+1) = [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)]^T \boldsymbol{Q} \, [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)] + \boldsymbol{u}^T(k)\boldsymbol{R}\boldsymbol{u}(k) \qquad (5)$$

This expression is minimized with respect to $\boldsymbol{u}(k)$ to obtain the following control law

$$\boldsymbol{u}(k) = -\left(\boldsymbol{B}^T(k)\boldsymbol{Q}\boldsymbol{B}(k) + \boldsymbol{R}\right)^{-1}\boldsymbol{B}^T(k)\boldsymbol{Q}\,[\boldsymbol{x}(k) - \boldsymbol{x}_D(k+1)] \qquad (6)$$

The weighting matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ allow the user to place more or less emphasis on the feature error and the control input. Their selection effects the response and stability of the tracking system. The $\boldsymbol{Q}$ matrix must be positive definite, and $\boldsymbol{R}$ must be positive semi-definite for a bounded response. Although no standard procedure exists for choosing the elements of $\boldsymbol{Q}$ and $\boldsymbol{R}$, general guidelines can be found in Papanikolopoulos *et al*. (1992).

2.3. Measurement of feature positions

The measurement of the motion of the features on the image plane must be done continuously and quickly. The method used to measure this motion is based on optical flow techniques and is a modification of the method proposed in Anandan (1987). This technique is known as Sum-of-Squares-Differences (SSD) optical flow, and is based on the assumption that the intensities around a feature point remain constant as that point moves across the image plane. A more complete description of the algorithm and its implementation can be found in Papanikolopoulos *et al*. (1992).

3. DYNAMIC SENSOR PLACEMENT CRITERIA

Several different criteria can be used to influence camera motion. This section presents three dynamically determined sensor placement criteria, as well as a technique for effectively integrating these criteria into the visual tracking control law.

3.1. Measure of focus

Keeping features in focus is important to the success of the SSD optical flow algorithm. Krotkov (1987) investigated several techniques for measuring the sharpness of focus. One problem with traditional focus measures is that they are dependent on the scale of the feature. When adjusting the focal ring, feature size changes only slightly so scaling effects can be ignored. However, in dynamic sensor placement strategies, changing the depth by moving the camera is the only way to bring a fixed focal length camera-lens system into focus. This means that greater changes in scale must be tolerated, and thus, traditional focal measures prove inadequate. Because of this, a Fourier transform based focal measure is used which analyzes the high frequency content of the feature in the most recent image to determine whether the object being tracked is within the depth-of-field of the camera.

3.2. Spatial resolution

A spatial resolution constraint is necessary for ensuring that parts being visually servoed by manipulators can be brought near enough their goal so that final mating can be successfully accomplished by force control. To incorporate spatial resolution constraints into the eye-in-hand system, it is assumed that maximum spatial resolution is always desired. Thus, the depth of the object from the camera, $Z_o(k)$, is to be minimized. It is also assumed that other sensor placement criteria, such as focus or field-of-view, will keep the depth from exceeding extreme values.

3.3. Manipulator configuration

A significant problem with eye-in-hand systems is the avoidance of kinematic singularities and joint limits. In Nelson and Khosla (1993), an efficient technique for avoiding singularities and joint limits while visually tracking is presented. A manipulability measure of the form

$$w(\boldsymbol{q}) = \left(1 - e^{-k\prod_{i=1}^{n}\frac{(q_i - q_{imin})\,(q_{imax} - q_i)}{(q_{imax} - q_{imin})^2}}\right)(det(\boldsymbol{J}(\boldsymbol{q})))^2 \tag{7}$$

is used to avoid singularities along redundant or unconstrained tracking axes. In (7), k is a user defined constant, n is the number of joints, q_i is the ith joint angle, q_{imin} and q_{imax} are the minimum and maximum allowable joint values, respectively, for the ith joint, and $\boldsymbol{J}(\boldsymbol{q})$ is the Jacobian matrix of a non-redundant manipulator. This measure is a modification of one proposed by Tsai (1986), which multiplies Yoshikawa's (1985) measure of nearness to singularities by a penalty function which indicates the distance to the nearest joint limit.

3.4. Augmenting the controller objective function

An effective technique for introducing sensor placement criteria into the eye-in-hand visual tracking control law is to include the placement measures in the controller objective function. To include the measures previously discussed, the objective function given by (5) becomes

$$F(k+1) = [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)]^T \boldsymbol{Q}\, [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)] + \boldsymbol{u}^T(k)\boldsymbol{R}\boldsymbol{u}(k) + \frac{S}{f_m} + UZ_o^2(k+1) + \frac{V}{w(\boldsymbol{q})} \tag{8}$$

The control law is derived by minimizing the objective function with respect to the control input, $\boldsymbol{u}(k)$. This results in a control law of the form

$$\boldsymbol{u}(k) = -\left(\boldsymbol{B}^T(k)\,\boldsymbol{Q}\boldsymbol{B}(k) + \boldsymbol{R}\right)^{-1} \times \left[\boldsymbol{B}^T(k)\,\boldsymbol{Q}\,[\boldsymbol{x}(k) - \boldsymbol{x}_D(k+1)] - \frac{S}{2f_m^2}\nabla^T_{\boldsymbol{u}(k)} f_m + UZ_o(k)T\begin{bmatrix}0 & 0 & 1 & 0 & 0 & 0\end{bmatrix}^T - \frac{V}{2w^2(\boldsymbol{q})}\nabla^T_{\boldsymbol{u}(k)} w(\boldsymbol{q})\right] \tag{9}$$

The terms representing the gradients of the focus measure f_m and manipulability w are positive semi-definite, can be approximated numerically in order to determine the current camera velocity which maximally increases their values. Relative weights are placed on the different criteria functions by S, U, and V. If the cartesian axes along which visual tracking takes place are different from axes along which any sensor placement criteria may influence, it is necessary to slightly alter $\boldsymbol{B_F}$ given in (2) in order to properly use this control law. The columns of $\boldsymbol{B_F}$ which correspond to cartesian axes along which visual tracking should not occur are simply set to zero. This inhibits visual tracking along these axes, but allows the cost term $\boldsymbol{u}^T\boldsymbol{R}\boldsymbol{u}$ to influence sensor motion along non-tracking axes due to dynamic sensor placement criteria. This also ensures that a six dimensional control input results from (9).

4. HARDWARE IMPLEMENTATION

The visual tracking algorithm described previously has been implemented on a robotic assembly system consisting of three Puma560's called the Rapid Assembly System. One of the Pumas has a Sony XC-77RR camera mounted at its end-effector. The camera is connected to an Innovision IDAS/150 Vision System. The Pumas are controlled from a VME bus with two Ironics IV-3230 (68030 CPU) processors, an IV-3220 (68020 CPU) which also communicates with a trackball, a Mercury floating point processor, and a Xycom parallel I/O board communicating with three Lord force sensors mounted on the Pumas' wrists. All processors on the controller VME run the Chimera real-time operating system (Stewart *et al*. 1993).

5. EXPERIMENTAL RESULTS

The experimental results were obtained with a strategy using motion about the three camera axes for visual tracking, motion along the three axes for singularity/joint limit avoidance, and motion along the camera's optical axis (Z) for increasing spatial resolution and maintaining focus. Two features on the target object were tracked so that the position and orientation of the object on the image plane could be maintained as the target moved, subject to the other sensor placement criteria. The object being tracked moved 25cm in a direction parallel to the camera's X axis. Without sensor placement criteria, the depth of the object should increase as illustrated by the dashed line in Fig. 2. However, the spatial resolution constraint causes the depth to actually decrease, until the decreased focal measure causes the camera to increase the depth in order to improve focus, as shown in Fig. 3.Without sensor placement criteria affecting depth and focus, the object would have become unfocused after moving less than 10cm. Instead, object motion of 25cm is successfully tracked.

Another criterion which, in this instance, causes the depth to increase is the singularity/joint limit avoidance criterion. Figure 4 shows that the manipulability measure decreases due to camera translation caused by the spatial resolution constraint. The decrease in manipulability occurs because the second and third joints of the Puma560 become nearly aligned, and the first joint nears a joint limit.

6. CONCLUSION

The working region of a camera providing visual feedback for robotic manipulation tasks can be significantly increased by combining visual tracking capabilities with dynamic sensor placement criteria. The controlled active vision paradigm provides a useful framework for integrating different sensor placement criteria into an eye-in-hand tracking system's control law. A system which accounts for focus, spatial resolution, and manipulator configuration has been experimentally verified. Future work will include a greater number of dynamic sensor placement criteria in the control law, such as field-of-view and occlusion avoidance, and will allow the observed object to be visually servoed by a second manipulator for automatic assembly.

7. REFERENCES

Anandan, P. (1987), "Measuring Visual Motion from Image Sequences," Technical Report COINS-TR-87-21, COINS Department, University of Massachusetts.

Cowan, C.K. and A. Bergman (1989), "Determining the Camera and Light Source Location for a Visual Task," In: *Proc. of the 1989 IEEE Int. Conf. on Robotics and Automation*, 509-514.

Nelson, B. and P.K. Khosla (1993), "Increasing the Tracking Region of an Eye-in-Hand System by Singularity and Joint Limit Avoidance," In: *Proc. of the 1993 IEEE Int. Conf. on Robotics and Automation*.

Papanikolopoulos, N.P., P.K. Khosla, and T. Kanade (1991), "Adaptive Robotic Visual Tracking," In: *Proc. of the American Control Conf.*, 962-967.

Papanikolopoulos, N.P., B. Nelson, and P.K. Khosla (1992),"Monocular 3-D Visual Tracking of a Moving Target by an Eye-in-Hand Robotic System, In: *Proc. of the 31st IEEE Conf. on Decision and Control (31stCDC)*, 3805-3810.

Stewart, D.B., D.E. Schmitz, and P.K. Khosla (1992), "The Chimera II Real-Time Operating System for Advanced Sensor-Based Control System," In: *IEEE Trans. Sys., Man and Cyb.*, **22**, 1282-1295.

Tarabanis, K., R.Y. Tsai, and P.K. Allen (1991), "Automated Sensor Planning for Robotic Vision Tasks," In: *Proc. of the 1991 IEEE Int. Conf. on Robotics and Automation*, 76-82.

Tsai, M.J. (1986), *Workspace Geometric Characterization of Industrial Robot*, Ph.D. Thesis, Ohio State University, Department of Mechanical Engineering.

Yi, S., R.M. Haralick, and L.G. Shapiro (1990),"Automatic Sensor and Light Positioning for Machine Vision," In: *Proc. of the 10th Int. Conf. on Pattern Recognition*, 55-59.

Yoshikawa, T. (1985), "Manipulability of Robotic Mechanisms," In: *Robotics Research 2*, eds. H. Hanafusa and H. Inoue, 439-446, MIT Press, Cambridge, MA.

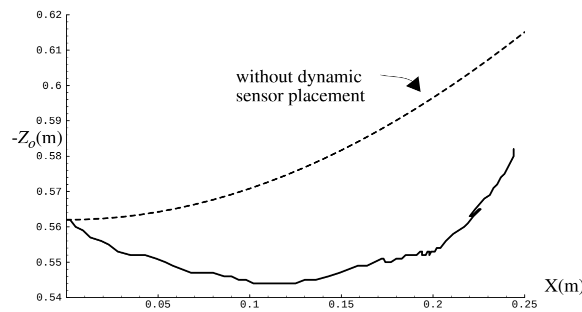

Fig. 2. Depth of object from camera frame origin versus distance object translates along X.

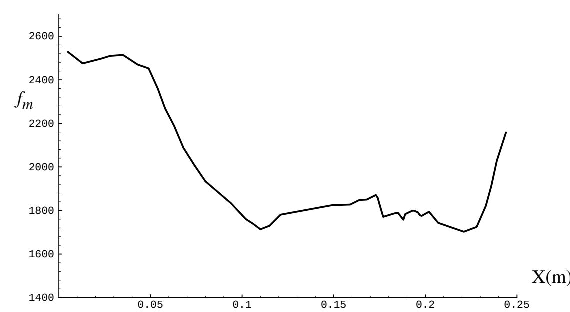

Fig. 3. Focal measure versus X translation of object.

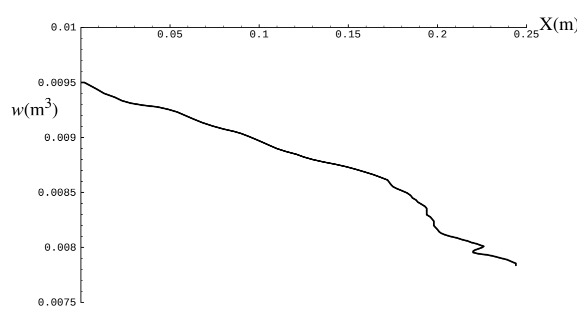

Fig. 4. Manipulability versus X translation of object.