EXPERIMENTAL ROBOTICS III:
The Third International Symposium,
Kyoto, October 28-30, 1993,
T. Yoshikawa and F. Miyazaki (Eds).

Integrating Sensor Placement and Visual Tracking Strategies

Brad Nelson Pradeep K. Khosla

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213-3891, USA

Abstract—Real-time visual feedback is an important capability that many robotic systems must possess if these systems are to operate successfully in dynamically varying and/or uncalibrated environments. An eye-in-hand system is a common technique for providing camera motion to increase the working region of a visual sensor. Although eye-in-hand robotic systems have been well-studied, several deficiencies in proposed systems make them inadequate for actual use. Typically, the systems fail if manipulators pass through singularities or joint limits. Objects being tracked can be lost if the objects become defocused, occluded, or if features on the objects lie outside the field of view of the camera. In this paper, a technique is introduced for integrating a visual tracking strategy with dynamically determined sensor placement criteria. This allows the system to automatically determine, in real-time, proper camera motion for tracking objects successfully while accounting for the undesirable, but often unavoidable, characteristics of camera-lens and manipulator systems. The sensor placement criteria considered include focus, field-of-view, spatial resolution, manipulator configuration, and a newly introduced measure called *resolvability.* Experimental results are presented.

1. Introduction

Real-time visual feedback is an important capability that many robotic systems must possess if these systems are to operate successfully in dynamically varying and/or uncalibrated environments. In order to significantly increase the working region of a sensor providing real-time visual feedback, it is necessary to allow the visual sensor to move. An eye-in-hand system is a common technique for providing camera motion. Although eye-in-hand robotic systems have been well-studied, several deficiencies in proposed systems make them inadequate for actual use. Tracking regions must be severely constrained so that manipulators do not pass through kinematic singularities or joint limits. Objects whose depth from the camera might significantly vary can become defocused to such an extent that the object is lost, or the projections of all features on the object of interest cannot be constrained to fall on the image plane simultaneously. Camera spatial resolution may not allow a sufficiently high degree of tracking accuracy. Past eye-in-hand systems have only been able to deal with these situations by severely constraining object motion to ensure that these problems cannot arise.

We propose a tracking strategy that allows hand/eye systems to operate successfully without encountering many of the problems that previous eye-in-hand tracking

systems have failed to address. Several factors are considered when controlling camera motion. Object motion induces appropriate camera motion to track the object of interest. If 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 task being performed in order to increase the spatial resolution of the sensor 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. All of these factors are allowed to potentially affect camera motion.

Camera motion can be induced by teleoperator input, as well. This allows a remote user looking only at the image produced by the camera to control camera motion without being concerned with manipulator singularities, joint limits, poorly focused objects, objects leaving the field of view of the camera, and poor spatial resolution. Systems of this type will prove useful for visually inspecting hazardous environments and for directing manipulation tasks being performed by other robots within these environments.

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 off-line ([2], [9], and [11]). 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, and sensor placement criteria must be integrated into the tracking strategy. The configuration of the manipulator must also be taken into account in the control strategy.

A previously proposed visual servoing paradigm [6] is used as a 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. 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. Our method of introducing different criteria into the control law results in hand/eye systems that can be programmed in an evolutionary way so that different behaviors can be easily introduced into the hand/eye controller. A brief description of the experimental system and presentation of experimental results complete the paper.

2. Modeling and Control of the Tracking System

We first present a system model and controller for visually tracking an object without considering any sensor placement criteria other than tracking the object. In Section 3, we will show how sensor placement criteria can be integrated into the system controller.

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 Figure 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

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

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 $\boldsymbol{B_F}(k)\in R^{2x6}$ is

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

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 non-collinear features are required [7]. 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

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

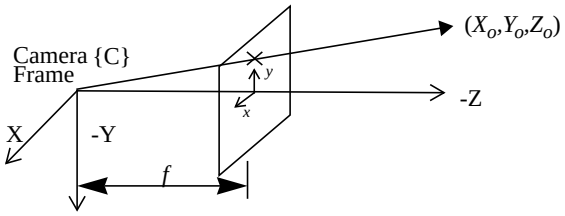

Figure 1. The pinhole camera model.

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

$$\boldsymbol{B}(k) = \begin{bmatrix} \boldsymbol{B}_{\boldsymbol{F}}^{(1)}(k) \\ \cdots \\ \boldsymbol{B}_{\boldsymbol{F}}^{(M)}(k) \end{bmatrix} \tag{4}$$

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 $\boldsymbol{x}(k)=[x^{(1)}(k)\ y^{(1)}(k)...\ x^{(M)}(k)\ y^{(M)}(k)]^{\mathrm{T}}$ is the new state vector, and $\boldsymbol{d}(k)$ is the new exogenous deterministic disturbances vector.

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

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

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

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

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 [6].

In Section 3, this basic controller formulation will be extended to include additional objectives rather than just feature tracking. These additional objectives create behaviors which cause the system to improve the spatial resolution of the visual data provided and to avoid undesirable camera-lens and manipulator characteristics. By introducing the objectives into the system through the control objective function, different behaviors can be easily tested to ensure the desired collective system response is achieved. Another advantage of introducing sensor placement objectives into the controller in this manor, is that the hand/eye system parameters directly affect the magnitude of the control response to the different criteria. Thus, each behavior's effect on the overall system response can be determined independent of the particular hand/eye system.

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 [1]. 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 [6].

3. Dynamic Sensor Placement Criteria

Several different criteria can be used to influence camera motion. This section presents several dynamically determined sensor placement criteria, as well as a technique for effectively integrating all of the 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. Several techniques for measuring the sharpness of focus are investigated in [3]. 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 focus 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.

A well accepted model of the point spread function of a camera-lens system is represented by a Gaussian distribution. Since defocusing corresponds to greater attenuation of high frequencies, the Gaussian distribution representing the point spread function of a camera-lens model becomes wider as the sharpness of focus decreases, and the Fourier transform of the camera lens system, which is also Gaussian, becomes narrower. One would therefore assume that the high frequency magnitudes of the Fourier transform of the feature window would become smaller. While this is true, it is also true that thresholds at which these high frequency magnitudes become small enough to indicate defocused features is dependent on the feature as well. This makes it difficult to determine whether the feature is significantly defocused, or whether the feature actually contains relatively few high frequency components.

In computing the fast Fourier transform of the 16x16 feature, windowing effects must be considered. The window introduces certain effects in the Fourier transform which, for a step edge, causes the Fourier transform to exhibit sinc-like behavior, as shown in Figure 2. As the sharpness of focus decreases, a Gaussian point spread function indicates that the Fourier transform becomes narrower as high frequencies become attenuated. Windowing effects become less pronounced and the Fourier transform looks less like a sinc function, as illustrated by Figure 3.

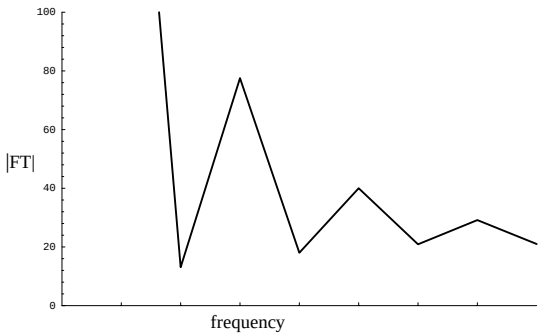

Figure 2. FT magnitude of a windowed step edge in sharp focus.

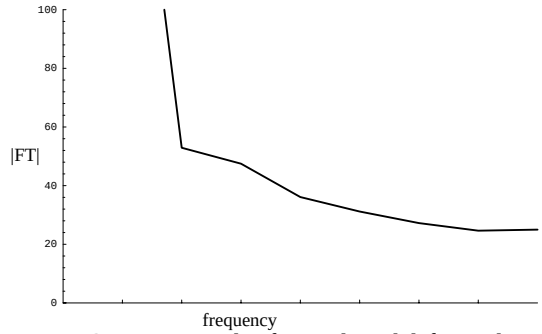

Figure 3. FT magnitude of a windowed defocused step edge.

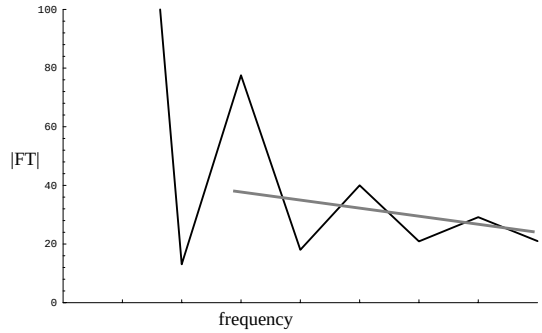

Figure 4. Least-squares line fit of FT high frequency magnitudes.

We propose to measure the degree of focus by observing how closely the high frequency components of the feature's Fourier transform approximate a sinc function. When features are sharply focused, a least-squares line fit of the high frequency components results in a large residual error shown by Figure 4. As features become less focused, the residual error decreases as the high frequency components fit a line well. Figure 5 shows the measure of focus as the object depth varies from 20cm to 80cm. At 46cm, the feature is in sharpest focus and the measure is very high. At 20cm, the linear scale of the feature is approximately 1.7 times the original scale, and at 80cm

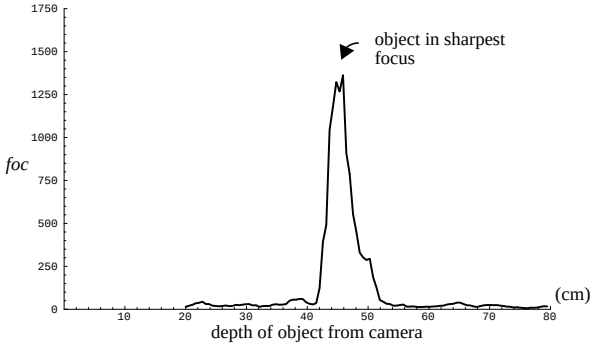

Figure 5. Proposed measure of focus of a feature on an object whose depth varies 60cm.

the scale is approximately 0.43 the original scale, but the measure is clearly independent of the overall scale change factor of four. Experimental results indicate that this measure is quite robust.

This measure can be introduced into the control objective function as

$$F(k+1) = [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)]^T \boldsymbol{Q} [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)] + \boldsymbol{u}^T(k)\boldsymbol{R}\boldsymbol{u}(k) + \frac{U}{(foc)} \quad (7)$$

resulting in a control law of the form

$$\boldsymbol{u}(k) = -\left(\boldsymbol{B}^T(k)\boldsymbol{Q}\boldsymbol{B}(k) + \boldsymbol{R}\right)^{-1}\left[\boldsymbol{B}^T(k)\boldsymbol{Q}[\boldsymbol{x}(k) - \boldsymbol{x}_D(k+1)] - \frac{U}{2(foc)^2}\nabla^T_{\boldsymbol{u}(k)}(foc)\right] \quad (8)$$

3.2. Spatial Resolution

A spatial resolution constraint is necessary for ensuring that objects are being observed with the maximum possible spatial resolution. For autonomous visually guided manipulation, this also allows parts being visually servoed to 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. Introducing the spatial resolution constraint into the objective function results in an objective function and control law of

$$F(k+1) = [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)]^T \boldsymbol{Q} [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)] + \boldsymbol{u}^T(k)\boldsymbol{R}\boldsymbol{u}(k) + VZ_o^2(k+1) \quad (9)$$

$$\boldsymbol{u}(k) = -\left(\boldsymbol{B}^T(k)\boldsymbol{Q}\boldsymbol{B}(k) + \boldsymbol{R}\right)^{-1} \times \left[\boldsymbol{B}^T(k)\boldsymbol{Q}[\boldsymbol{x}(k) - \boldsymbol{x}_D(k+1)] + VZ_o(k)T\begin{bmatrix}0 & 0 & 1 & 0 & 0 & 0\end{bmatrix}\right] \quad (10)$$

3.3. Field of View

In order to ensure that the projections of features being observed do not exceed the boundaries of the image plane causing the features to be lost, it is necessary to introduce a field of view constraint. A potential function can be created on the image

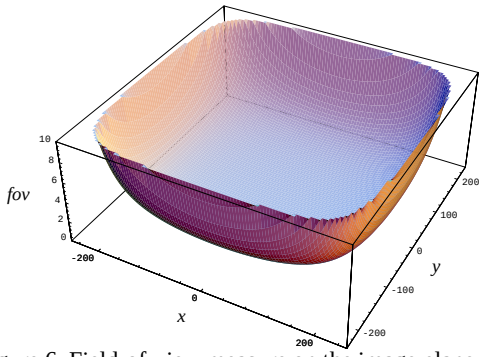

Figure 6. Field-of-view measure on the image plane.

plane, which maintains a constant value away from the image plane boundaries and approaches infinity at the boundaries. The function can be represented analytically as

$$fov = \prod_{i=1}^{n} \left(1 - \frac{x_i^2(k+1)}{x_{Bi}^2} \right) \tag{11}$$

where n is the number of system states, and x_i are the states of the system, which is equivalent to the feature coordinates. The term x_{Bi} represents the bound on the absolute value of the particular state, which is either the maximum x or y coordinate on the image plane.

Figure 6 illustrates the potential function graphically. As camera or object motion causes feature projections to approach the edges of the CCD array, the gradient of this function in cartesian space can be analytically calculated, and appropriate camera motion directions determined to cause the feature projections to move away from the boundaries.

The field of view measure can be introduced into the control objective function as

$$F(k+1) = [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)]^T \boldsymbol{Q} [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)] + \boldsymbol{u}^T(k) \boldsymbol{R} \boldsymbol{u}(k) + \frac{W}{(fov)} \tag{12}$$

resulting in a control law of the form

$$\boldsymbol{u}(k) = -\left(\boldsymbol{B}^T(k) \boldsymbol{Q} \boldsymbol{B}(k) + \boldsymbol{R} \right)^{-1} \times \left[\boldsymbol{B}^T(k) \boldsymbol{Q} [\boldsymbol{x}(k) - \boldsymbol{x}_D(k+1)] - \frac{W}{2(fov)^2} \nabla^T_{\boldsymbol{u}(k)} (fov) \right] \tag{13}$$

3.4. Manipulator Configuration

A significant problem with eye-in-hand systems is the avoidance of kinematic singularities and joint limits. In [4], an efficient technique for avoiding singularities and joint limits while visually tracking is presented. A manipulability measure of the form

$$w(\boldsymbol{q}) = \left(1 - e^{-k \prod_{i=1}^{n} \frac{(q_i - q_{imin})(q_{imax} - q_i)}{(q_{imax} - q_{imin})^2}} \right) (det(\boldsymbol{J}(\boldsymbol{q})))^2 \tag{14}$$

is used to avoid singularities along redundant or unconstrained tracking axes. In (14), 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 [10], which multiplies Yoshikawa's [12] measure of nearness to singularities by a penalty function which indicates the distance to the nearest joint limit. In [4], experimental results are presented which show that tracking regions of hand/eye systems can be significantly extended by incorporating singularity and joint limit avoidance into the tracking strategy.

The objective function and control law for including the manipulability measure is represented by

$$F(k+1) = [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)]^T \boldsymbol{Q} [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)] + \boldsymbol{u}^T(k)\boldsymbol{R}\boldsymbol{u}(k) + \frac{S}{w'(\boldsymbol{q}(k))} \quad (15)$$

$$\boldsymbol{u}(k) = -\left(\boldsymbol{B}^T(k)\boldsymbol{Q}\boldsymbol{B}(k) + \boldsymbol{R}\right)^{-1}\left[\boldsymbol{B}^T(k)\boldsymbol{Q}[\boldsymbol{x}(k) - \boldsymbol{x}_D(k+1)] - \frac{S}{2w'(\boldsymbol{q})^2}\nabla^T_{\boldsymbol{u}(k)} w'(\boldsymbol{q})\right] \quad (16)$$

3.5. Augmenting the Controller Function for Achieving Multiple Objectives

Some of the previously proposed placement measures can result in systems that do not function properly when implemented individually. For example, the spatial resolution constraint would tend to drive the camera toward the object to reduce the depth to zero. It becomes necessary to include a focus constraint or a field-of-view constraint to ensure that the depth does not become too small. An advantage of the controller formulation we propose is that all of these measures can be easily combined into a system that attempts to satisfy all system constraints collectively. To include all of the positive semi-definite measures previously discussed into a single control law, the objective function given by (5) becomes

$$F(k+1) = [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)]^T \boldsymbol{Q} [\boldsymbol{x}(k+1) - \boldsymbol{x}_D(k+1)] + \\ + \boldsymbol{u}^T(k)\boldsymbol{R}\boldsymbol{u}(k) + \frac{S}{w'(\boldsymbol{q}(k))} + \frac{U}{(foc)} + VZ_o^2(k+1) + \frac{W}{(fov)} \quad (17)$$

This results in a control law of the form

$$\boldsymbol{u}(k) = -\left(\boldsymbol{B}^T(k)\boldsymbol{Q}\boldsymbol{B}(k) + \boldsymbol{R}\right)^{-1}\left[\boldsymbol{B}^T(k)\boldsymbol{Q}[\boldsymbol{x}(k) - \boldsymbol{x}_D(k+1)] - \frac{S}{2w'(\boldsymbol{q})^2}\nabla^T_{\boldsymbol{u}(k)} w'(\boldsymbol{q}) - \right. \\ \left. - \frac{U}{2(foc)^2}\nabla^T_{\boldsymbol{u}(k)}(foc) + VZ_o(k)T\begin{bmatrix}0 & 0 & 1 & 0 & 0 & 0\end{bmatrix} - \frac{W}{2(fov)^2}\nabla^T_{\boldsymbol{u}(k)}(fov)\right] \quad (18)$$

The terms representing the gradients of the focus measure *foc* and manipulability *w* can be approximated numerically in order to determine the current camera velocity which maximally increases their values. The field-of-view measure *fov* can be calculated analytically. Relative weights are placed on the different criteria functions by *S*, *U*, *V*, and *W*. 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 (18).

3.6. *Resolvability*

When an eye-in-hand system is used to provide visual input for a second manipulator performing a manipulation task, the sensor placement measure *resolvability* can be used in place of the spatial resolution constraint presented in Section 3.2. This measure provides a technique for estimating the relative ability of various visual sensor systems, including single camera systems, stereo pairs, multi-baseline stereo systems, and 3D rangefinders, to accurately control visually manipulated objects.

The term *resolvability* refers to the ability of a visual sensor to resolve object positions and orientations. The *resolvability ellipsoid* [5] illustrates the directional nature of *resolvability*, and can be used to direct camera motion and adjust camera intrinsic parameters in real-time so that the servoing accuracy of the visual servoing system improves with camera-lens motion. The object centered Jacobian mapping from task space to sensor space is an essential component of the sensor placement measure. A singular value decomposition of this mapping provides the six-dimensional *resolvability* measure, which can be interpreted as the system's ability to resolve task space positions and orientations on the sensor's image plane.

Figure 7 shows the *resolvability* in depth of an object 10cm in length lying in a plane parallel to the image plane. As one might expect, the plot shows that when considering the limited size of the camera CCD it is preferable to decrease depth as much as possible in order to increase *resolvability* in depth, rather than to increase the focal length of the lens. This directs the system to move the camera closer to the object rather than to increase the focal length of the lens. However, this measure does not account for the depth-of-field of the lens. When the focus measure presented in Section 3.1 is considered, the depth of the object cannot decrease beyond a particular value, depending on lens properties, before the object becomes severely defocused and untrackable. Thus, a tradeoff between focal length, depth, and depth-of-field must be achieved when considering *resolvability* in depth. Our proposed control strategy which allows multiple camera-lens-manipulator behaviors to influence camera-lens motion allows this tradeoff to be achieved. Visual servoing experiments demonstrate that *resolvability* can be successfully used to direct camera-lens motion in order to increase the ability of a visually servoed manipulator to precisely servo objects [5].

4. Experimental Results

The visual tracking algorithm described previously has been implemented on a robotic assembly system consisting of three Puma 560'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 a Datacube Maxtower 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 [8].

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 Figure 8. 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 Figure 9. Without sensor placement criteria affecting depth and focus, the object would have become defocused to such an extent

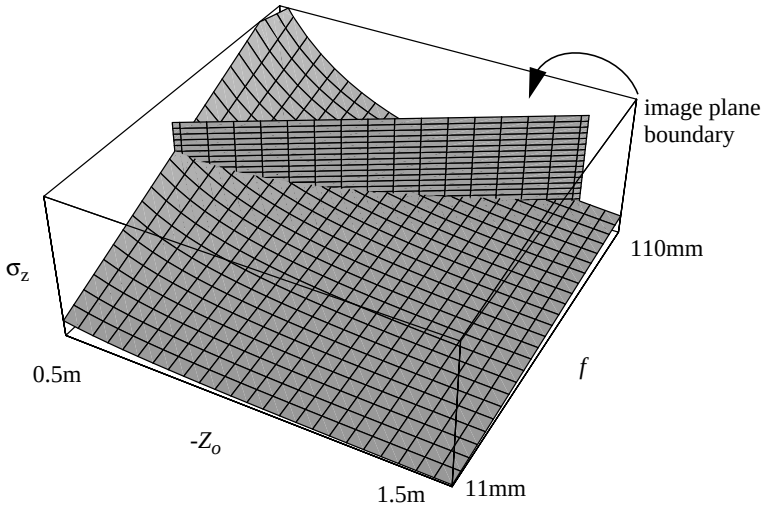

Figure 7. *Resolvability* of depth σ_z versus depth of object and focal length for two features 10cm apart lying on an object in a plane parallel to the image plane.

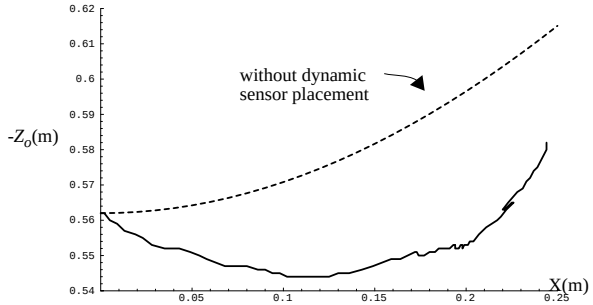

Figure 8. Depth of object from camera frame origin versus distance object translates along X.

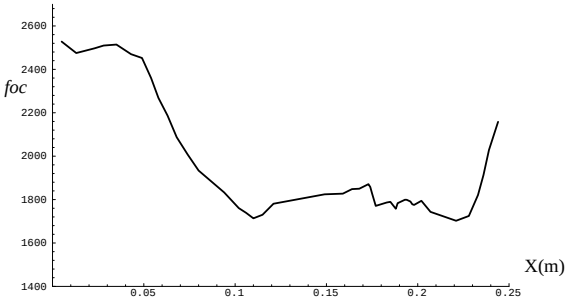

Figure 9. Focal measure versus X translation of object.

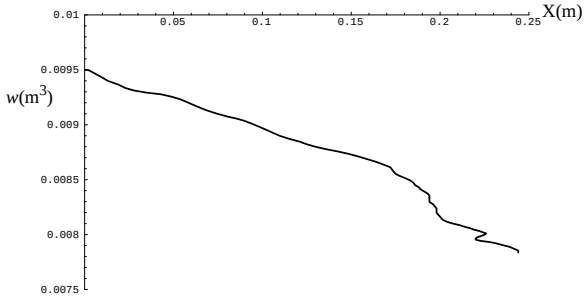

Figure 10. Manipulability versus X translation of object.

that the SSD tracker would have failed after the object moved 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 10 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 Puma 560 become nearly aligned, and the first joint nears a joint limit.

5. 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 tracking 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.

This tracking strategy is also useful when teleoperator input based on image observation alone is provided to a hand/eye system remotely. The operator is relieved of the need to concern him or herself with the possibility of causing the manipulator to pass through a kinematic singularity or joint limit. Recent experiments have successfully demonstrated the capability to teleoperate a hand/eye system at Carnegie Mellon's Advanced Manipulators Lab in Pittsburgh, Pennsylvania from Sandia National Laboratories in Albuquerque, New Mexico. The teleoperator in Albuquerque inspected an object in Pittsburgh by moving the camera mounted on a Puma 560 around the object

while the system simultaneously visually tracked the object. The operator's feedback (in Albuquerque) came solely from images relayed from the hand/eye camera (in Pittsburgh). Although delays of up to one second sometimes occurred during the image transfer, the hand/eye system at Carnegie Mellon continued to operate successfully and provided useful visual input to the teleoperator without entering singular configurations or violating joint limits.

Future work will include a greater number of dynamic sensor placement criteria in the control law, such as occlusion avoidance, and will allow the observed object to be visually servoed by a second manipulator for automatic assembly while using the newly introduced *resolvability* sensor placement measure. A motorized zoom/focus/aperture lens has recently been incorporated into the system in order to further enhance system capabilities, and control issues related to dynamically reconfigurable sensors are currently being explored.

Acknowledgments

This research was supported by the U.S. Army Research Office through Grant Number DAAL03-91-G-0272 and by Sandia National Laboratories through Contract Number AC-3752D.

References

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

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

[3] E. Krotkov, "Focusing," *International Journal of Computer Vision*, (1), 223-237, 1987.

[4] B. Nelson and P.K. Khosla, "Increasing the Tracking Region of an Eye-in-Hand System by Singularity and Joint Limit Avoidance," *Proc. of the 1993 IEEE Int. Conf. on Robotics and Automation*, 3:418-3:423, 1993.

[5] B. Nelson and P.K. Khosla, "The Resolvability Ellipsoid for Visual Servoing," Technical Report CMU-RI-TR-93-28, The Robotics Institute, Carnegie Mellon University., 1993.

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

[7] N.P. Papanikolopoulos and P.K. Khosla, "Robotic Visual Servoing Around a Static Target: An Example of Controlled Active Vision," *Proc. of the 1992 American Control Conference*,1489-1494, 1992.

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

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

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

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

[12] T. Yoshikawa, "Manipulability of Robotic Mechanisms," *Robotics Research 2*, eds. H. Hanafusa and H. Inoue, 439-446, MIT Press, Cambridge, MA, 1985.