Presented at The 34th IEEE Conference on Decision and Control, New Orleans, LA, Dec. 13-15, 1995.

Theory and Experiments in Vision-Based Grasping

Christopher E. Smith Nikolaos P. Papanikolopoulos
Artificial Intelligence, Robotics, and Vision Laboratory
Department of Computer Science
University of Minnesota
4-192 EE/CS Building
200 Union St. SE
Minneapolis, MN 55455

Abstract

Flexible operation of a robotic agent requires interaction with an uncalibrated or partially calibrated environment through the use of sensing. Much of the recent work in robotics and computer vision has concentrated upon the active observation of dynamic targets by the robotic agent. This paper focuses on autonomous interaction with moving targets in the environment. In particular, we propose a system that performs autonomous grasping of a moving target in an uncalibrated environment. The proposed system is derived using the Controlled Active Vision framework and provides the flexibility to robustly interact with the environment in the presence of uncertainty. The proposed work is experimentally verified using the Minnesota Robotic Visual Tracker (MRVT) to select targets of interest, to derive estimates of unknown environmental parameters, and to supply a control vector based upon these estimates to guide the manipulator in both the tracking and the grasping of a target.

1. Introduction

Current industrial manipulators suffer from ineffectiveness due to their inability to perform satisfactorily in a variety of situations. Current systems are often very brittle and fail due to changes in the environment, the manipulator, or the sensors. Typically, objects to be manipulated are required to appear in distinguished positions and at pre-defined orientations (often through the aid of fixtures), or (if moving) are required to maintain stringent speed, location, and orientation restrictions. If these restrictions are not adhered to, then the system fails with no hope of recovery via sensing.

Flexible manipulation of objects requires the use of sensors in order to determine salient properties of the object of interest and the robot's workspace. The recent introduction of inexpensive and fast real-time image processing systems allows for the efficient integration of visual sensory information in the feedback loop of a robotic system. Even though the robotic visual control area has drastically expanded in the recent years, its main focus has remained the visual tracking of objects by using the information gathered by static- or robot-mounted cameras [2][6][12][17][18][20]. This work, while important in its results and implications, has concentrated upon the active observation of the environment, leaving interaction as an issue for future research. In particular, only a small number of researchers [1][8][9] [16] have proposed vision-based robotic systems that interact with the environment.

We propose a flexible system based upon a camera repositioning controller operating under the Controlled Active Vision framework [12][14]. The controller allows the manipulator to robustly grasp objects present in the workspace (see Figure 1). The system operates in an uncalibrated space with an uncalibrated camera. Moreover, the proposed scheme allows automatic planning and execution of all the necessary actions in order to grasp an object. The object of interest is not required to appear in a specific location, orientation, or depth, nor is it required to remain motionless during grasping.

In this paper, we first discuss the visual measurements we have used in this problem, elaborate on the use of "coarse" and "fine" features for guiding grasping, and discuss feature selection and reselection. We then describe the application of the Controlled Active Vision framework to the problem of vision-based grasping of objects. We verify the operation of the system by presenting experimental results using the MRVT [4] system where the manipulator successfully grasps moving objects using a vision-based, closed loop control strategy throughout the task. Finally, we discuss the strengths and weaknesses of our approach, suggest required future work, and summarize our results.

2. Vision-Based Control for Grasping

Measurements

We assume a pinhole camera model with a world frame $\{R_S\}$ fixed with respect to the camera and the Z-axis pointing along the optical axis. A point $\mathbf{P} = (X_S, Y_S, Z_S)^T$ in $\{R_S\}$ projects to a point $\mathbf{p}$ in the image plane with image coordinates x and y. For simplicity, we assume that $\delta_x = \delta_y = f = 1$, where δ_x and δ_y are the scaling factors for pixel size and camera sampling and f is the camera focal length.

By utilizing the derivation presented previously in [12][17][18], we arrive at the following equations describing the motion of $\mathbf{p}$ on the image plane due to $\mathbf{P}$ moving with translational motion $\mathbf{t} = (t_x, t_y, t_z)^T$ and

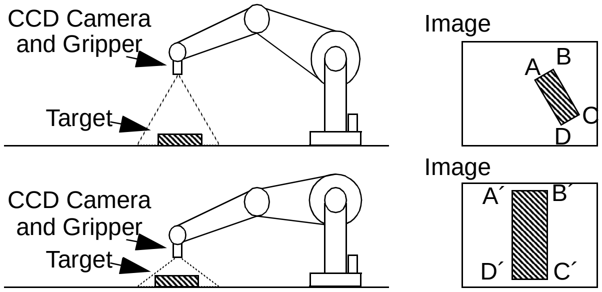

Figure 1: Experimental setup

rotational motion $\mathbf{r} = (r_x, r_y, r_z)^{\mathrm{T}}$:

$$u = \dot{x} = \left[x\frac{t_z}{Z_S} - \frac{t_x}{Z_S}\right] + [xyr_x - (1 + x^2)r_y + yr_z] \tag{2.1}$$

$$v = \dot{y} = \left[y\frac{t_z}{Z_S} - \frac{t_y}{Z_S}\right] + [(1 + y^2)r_x - xyr_y - xr_z] \,. \tag{2.2}$$

The continuous extraction of the positions of the features' projections on the image plane is based on optical flow techniques as presented in [12][17][19].

Coarse and fine features

We decompose the motion into coarse and fine segments by using two different classes of object features during operation as presented in [19]. Due to the wide range of relative object depth, initial object features will pass out of the view of the camera due to looming during the grasp reach. Therefore, we use "coarse" and "fine" features to guide manipulator. Coarse features are used to initially align the gripper and to begin the reach. When the coarse features approach the boundaries of the image plane, fine features are selected. These are used to drive the end-effector the remaining distance to the object and to signal when to grasp the object using a pneumatic, two-fingered hand. Proper orientation is maintained throughout by visual information derived from either the coarse or the fine features, depending upon the type of features being used to guide the manipulator.

Feature selection/reselection

An algorithm based upon the SSD technique may fail due to repeated patterns in the intensity function of the image or due to large areas of uniform intensity in the image. Both cases can provide multiple matches within a feature point's neighborhood, resulting in incorrect displacement measures. Furthermore, during certain movements of the manipulator (e.g., Z-axis translation and X-, Y-, or Z-axis rotations), the features being tracked will be distorted on the image plane, resulting in loss of tracking. In order to avoid these problems, our system automatically evaluates, selects, and reselects feature points as presented previously in [18][19].

Feature reselection is performed every μ th iteration in a small area Γ about each feature point using the method described in [18][19]. This prevents loss of feature tracking due to distortions caused by the rotation about the Z-axis required to align the gripper with the object and the motion of the object. The reselection rate μ is based upon the maximum rotation rate about Z, the estimated motion of the object, and the expected velocity of the feature points on the image plane.

3. Grasping as a visual servoing problem

We address the problem of grasping (eye-in-hand configuration) as a visual servoing problem in this section. The grasping problem can be defined as "find the motion of the manipulator that will grasp a static or slowly moving object." Since we are dealing with an eye-in-hand robotic system, we have to address the repositioning of the manipulator in order to effect grasping. The specific problem can be stated as "find the motion of the manipulator that will cause the image projections of certain feature points of the rigid target to move to desired image positions." We accomplish this by automatically defining desired positions for the object features such that the robot aligns the end-effector with the object, reaches toward the object (while maintaining gripper/object alignment), and grasps the object. Contrary to previous research efforts [5], only partial knowledge of the inverse perspective transformation is assumed.

Modeling approach

One feature point is not enough for the calculation of the control input vector due to the fact that the number of outputs is less than the number of inputs. Thus, we are obliged to consider more points in our model. In order to make the number of inputs equal to the number of outputs, we must consider at least three feature points which are not collinear. Having more than three feature points will result in a larger number of outputs than inputs. In grasping, the robot-camera system is required to take a certain pose with respect to the rigid target and this task requires at least four feature points.

According to the derivation given in [15], we produce the following equations written in the state-space form for four features (this model holds for static or slowly moving objects):

$$\begin{aligned}\mathbf{x}(k+1) = {} & \mathbf{A}(k)\mathbf{x}(k) + \\ & \mathbf{J}(k-d+1)\mathbf{u}_{con}(k-d+1) + \\ & \mathbf{H}(k)\mathbf{v}(k)\end{aligned} \tag{3.1}$$

where $\mathbf{A}(k) = \mathbf{H}(k) = \mathbf{I}_8$, $\mathbf{v}(k) \in \Re^8$, and d is the delay factor. The vector $\mathbf{u}_{con}(k) = (t_x(k), t_y(k), t_z(k), r_x(k), r_y(k), r_z(k))^{\mathrm{T}}$ is the control input vector. The matrix $\mathbf{J}(k) \in \Re^{8 \times 6}$ is (T is the sampling period):

$$\mathbf{J}(k) = \begin{bmatrix} \mathbf{J}_{\mathbf{F}}^{(1)}(k) \\ \mathbf{J}_{\mathbf{F}}^{(2)}(k) \\ \mathbf{J}_{\mathbf{F}}^{(3)}(k) \\ \mathbf{J}_{\mathbf{F}}^{(4)}(k) \end{bmatrix}$$

where each $\mathbf{J}_{\mathbf{F}}^{(i)}(k)$

$$\mathbf{J}_{\mathbf{F}}(k) =$$

$$T\begin{bmatrix} \dfrac{-1}{Z_s^{(i)}(k)} & 0 & \dfrac{x^{(i)}(k)}{Z_s^{(i)}(k)} & x^{(i)}(k)y^{(i)}(k) & -(1 + x^{(i)2}(k)) & y^{(i)}(k) \\ 0 & \dfrac{-1}{Z_s^{(i)}(k)} & \dfrac{y^{(i)}(k)}{Z_s^{(i)}(k)} & (1 + y^{(i)2}(k)) & -x^{(i)}(k)y^{(i)}(k) & -x^{(i)}(k) \end{bmatrix}.$$

The superscript (i) denotes each one of the feature points ($i \in \{1, 2, 3, 4\}$). The vector $\mathbf{x}(k) = (x^{(1)}(k), y^{(1)}(k), x^{(2)}(k), y^{(2)}(k), x^{(3)}(k), y^{(3)}(k), x^{(4)}(k), y^{(4)}(k))^{\mathrm{T}}$ is the state vector, and $\mathbf{v}(k) = (v_1^{(1)}(k), v_2^{(1)}(k), v_1^{(2)}(k), v_2^{(2)}(k), v_1^{(3)}(k), v_2^{(3)}(k), v_1^{(4)}(k), v_2^{(4)}(k))^{\mathrm{T}}$ is the white noise vector. The measurement vector

$$\mathbf{y}(k) = (y_1^{(1)}(k), y_2^{(1)}(k), y_1^{(2)}(k), y_2^{(2)}(k), y_1^{(3)}(k),$$

$y_2^{(3)}(k), y_1^{(4)}(k), y_2^{(4)}(k))^{\mathrm{T}}$ is given by:

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{w}(k) \qquad (3.2)$$

where $\mathbf{w}(k) = (w_1^{(1)}(k), w_2^{(1)}(k), w_1^{(2)}(k), w_2^{(2)}(k), w_1^{(3)}(k), w_2^{(3)}(k), w_1^{(4)}(k), w_2^{(4)}(k))^{\mathrm{T}}$ is the white noise vector $(\mathbf{w}(k) \sim N(\mathbf{0}, \mathbf{W}))$ and $\mathbf{C} = \mathbf{I}_8$. The measurement vector is computed using the SSD algorithm.

We can form a MIMO (Multi-Input Multi-Output) ARX (AutoRegressive with auXiliary input). This model consists of eight MISO (Multi-Input Single-Output) ARX models, and is described by the following equation:

$$\mathbf{A}(k)(1 - q^{-1})\mathbf{y}(k) = \mathbf{J}(k - d)\mathbf{u}_{con}(k - d) + \mathbf{n}(k) \qquad (3.3)$$

where $\mathbf{n}(k)$ is the white noise vector and q^{-1} is the backward shift operator. The white noise vector $\mathbf{n}(k)$ corresponds to the measurement noise, modeling errors, and noise introduced by inaccurate robot control. In the next section, we present the control and estimation techniques for the repositioning problem.

Control and estimation for repositioning

In order to grasp an object using an eye-in-hand system, the camera/manipulator must be repositioned with respect to the target. We use a repositioning scheme where the control objective is to move the manipulator in such a way that the projections of the selected features on the image plane move to some desired positions [13]. This section presents the control strategies that realize this motion and the estimation scheme used to estimate the unknown parameters of the model. Since the depth information is not directly available, adaptive control techniques are used for visually servoing around a object. In particular, adaptive control techniques are used for the recovery of the components of the translational and rotational velocity vectors $\mathbf{t}(k)$ and $\mathbf{r}(k)$, respectively. The rest of the section will be devoted to the description of the control and estimation schemes.

Control scheme for repositioning: The objective is to move the features' projections on the image plane to some desired positions. The repositioning of the projections is realized by an appropriate motion of the camera. The design of this controller is similar to the one proposed in [15]. By transforming our objective to a cost function, we can create a mathematical formula that continuously computes the desired motion of the camera. This motion is transformed through a robot control scheme to robot motion. In particular, a simple control law can be derived by the minimization of a cost function that includes the control signal [10]:

$$\begin{aligned} F(k + d) = {} & [\mathbf{y}(k + d) - \mathbf{y}_{des}(k + d)]^{\mathrm{T}}\mathbf{G_M} \\ & [\mathbf{y}(k + d) - \mathbf{y}_{des}(k + d)] + \\ & \mathbf{u}_{con}^{\mathrm{T}}(k)\mathbf{G_I}\mathbf{u}_{con}(k). \end{aligned} \qquad (3.4)$$

The vector $\mathbf{y}_{des}(k)$ represents the desired positions of the projections of the four features on the image plane. During certain stages of the grasping the vector $\mathbf{y}_{des}(k)$ is known but time-varying. By weighting the control signal, we place some emphasis on the minimization of the control signal in addition to the minimization of the servoing error. The response of the system is slower than having $\mathbf{G_I} = \mathbf{0}$ but the control input signal is bounded and feasible. This is in agreement with the structural and operational characteristics of the robotic system and the vision algorithm. A robotic system cannot track signals that command large changes in the features' image projections during the sampling interval *T*. The control law which is derived from the minimization of the cost function (3.4) is:

$$\begin{aligned} \mathbf{u}_{con}(k) = {} & -[\mathbf{J}^{\mathrm{T}}(k)\mathbf{G_M}\mathbf{J}(k) + \mathbf{G_I}]^{-1}\mathbf{J}^{\mathrm{T}}(k)\mathbf{G_M} \\ & \{\mathbf{y}(k) - \mathbf{y}_{des}(k + d) + \\ & \left.\sum_{m=1}^{d-1}\mathbf{J}(k - m)\mathbf{u}_{con}(k - m)\right\}. \end{aligned} \qquad (3.5)$$

The design parameters in this control law are the elements of the matrices $\mathbf{G_M}$ and $\mathbf{G_I}$. The matrix $\mathbf{G_M}$ should be positive definite ($\mathbf{G_M} > \mathbf{0}$) while $\mathbf{G_I}$ should be positive semidefinite ($\mathbf{G_I} \geq \mathbf{0}$). If the matrix $\mathbf{J}(k)$ is full rank then the matrix $\mathbf{J}^{\mathrm{T}}(k)\mathbf{G_M}\mathbf{J}(k) + \mathbf{G_I}$ is invertible. The matrix $\mathbf{J}(k)$ is singular when the four feature points are collinear. This is similar to the work presented in [15] that extends the number of points to *m*. For grasping, we will use four points. Further details on the conditions for singularity and a proof that those conditions make $\mathbf{J}(k)$ singular can be found in [12].

By selecting $\mathbf{G_I}$ and $\mathbf{G_M}$, one can place more or less emphasis on the control input and the servoing error. By following the results in [15], we can select the elements of these matrices. If we want to include the noise of our model and the inaccuracy of the $\mathbf{J}(k)$ matrix in our control law, the control objective (3.4) will become:

$$\begin{aligned} F(k + d) = {} & E\{[\mathbf{y}(k + d) - \mathbf{y}_{des}(k + d)]^{\mathrm{T}}\mathbf{G_M} \\ & [\mathbf{y}(k + d) - \mathbf{y}_{des}(k + d)] + \\ & \mathbf{u}^{\mathrm{T}}{}_{con}(k)\mathbf{G_I}\mathbf{u}_{con}(k) | F_k\} \end{aligned} \qquad (3.6)$$

where the symbol $E\{X\}$ denotes the expected value of the random variable *X* and F_k is the sigma algebra generated by the past measurements and the past control inputs up to time *k*. The new control law is:

$$\begin{aligned} \mathbf{u}_{con}(k) = {} & -[\hat{\mathbf{J}}^{\mathrm{T}}(k)\mathbf{G_M}\hat{\mathbf{J}}(k) + \mathbf{G_I}]^{-1}\hat{\mathbf{J}}^{\mathrm{T}}(k)\mathbf{G_M} \\ & \{[\mathbf{y}(k) - \mathbf{y}_{des}(k + d)] + \\ & \sum_{m=1}^{d-1}\hat{\mathbf{J}}(k - m)\mathbf{u}_{con}(k - m)\} \end{aligned} \qquad (3.7)$$

where $\hat{\mathbf{J}}(k)$ is the estimated value of the matrix $\mathbf{J}(k)$. The matrix $\hat{\mathbf{J}}(k)$ is dependent on the estimated values of the features' depth $\hat{Z}_s^{(i)}(k)$ $(i \in \{1, 2, 3, 4\})$ and the coordinates of the features' image projections. In particular, the matrix $\hat{\mathbf{J}}(k)$ is defined as follows:

$$\hat{\mathbf{J}}(k) = \begin{bmatrix} \hat{\mathbf{J}}_{\mathbf{F}}^{(1)}(k) \\ \hat{\mathbf{J}}_{\mathbf{F}}^{(2)}(k) \\ \hat{\mathbf{J}}_{\mathbf{F}}^{(3)}(k) \\ \hat{\mathbf{J}}_{\mathbf{F}}^{(4)}(k) \end{bmatrix}$$

where $\hat{\mathbf{J}}_{\mathbf{F}}^{(i)}(k)$ is given by:

$$\hat{\mathbf{J}}_{\mathbf{F}}^{(i)}(k) =$$

$$T\begin{bmatrix} \frac{-1}{\hat{Z}_s^{(i)}(k)} & 0 & \frac{x^{(i)}(k)}{\hat{Z}_s^{(i)}(k)} & x^{(i)}(k)y^{(i)}(k) & -[1+(x^{(i)}(k))^2] & y^{(i)}(k) \\ 0 & \frac{-1}{\hat{Z}_s^{(i)}(k)} & \frac{y^{(i)}(k)}{\hat{Z}_s^{(i)}(k)} & [1+(y^{(i)}(k))^2] & -x^{(i)}(k)y^{(i)}(k) & -x^{(i)}(k) \end{bmatrix}.$$

This matrix uses the estimated depth ($1/\hat{Z}_s^{(i)}(k)$) in the calculation of $\hat{\mathbf{J}}_{\mathbf{F}}^{(i)}(k)$. In the next section, we present estimation techniques for estimating the depth factor.

Computation of $\hat{\mathbf{J}}_{\mathbf{F}}^{(i)}(k)$ by estimating $1/Z_s^{(i)}(k)$: The estimation of the feature's depth $Z_s^{(i)}(k)$ with respect to the camera frame can be done in multiple ways. In this section, we present one estimation algorithm. Many more similar algorithms can be found in [15]. Let us define the inverse of the depth $Z_s^{(i)}(k)$ as $\zeta_s^{(i)}(k)$. Then, the equations of each feature point can be written as:

$$\begin{aligned} \mathbf{y}_{\mathbf{F}}^{(i)}(k) = {} & \mathbf{A}_{\mathbf{F}}^{(i)}(k-1)\mathbf{y}_{\mathbf{F}}^{(i)}(k-1) + \\ & \zeta_s^{(i)}(k-d)\mathbf{J}_{\mathbf{F1}}^{(i)}(k-d)\mathbf{t}(k-d) + \\ & \mathbf{J}_{\mathbf{F2}}^{(i)}(k-d)\mathbf{r}(k-d) + \mathbf{n}_{\mathbf{F}}^{(i)}(k) \end{aligned} \qquad (3.8)$$

where $\mathbf{J}_{\mathbf{F1}}^{(i)}(k)$ and $\mathbf{J}_{\mathbf{F2}}^{(i)}(k)$ are given by:

$$\mathbf{J}_{\mathbf{F1}}^{(i)}(k) = T\begin{bmatrix} -1 & 0 & x^{(i)}(k) \\ 0 & -1 & y^{(i)}(k) \end{bmatrix},$$

$$\mathbf{J}_{\mathbf{F2}}^{(i)}(k) =$$

$$T\begin{bmatrix} x^{(i)}(k)y^{(i)}(k) & -[1+(x^{(i)}(k))^2] & y^{(i)}(k) \\ [1+(y^{(i)}(k))^2] & -x^{(i)}(k)y^{(i)}(k) & -x^{(i)}(k) \end{bmatrix}.$$

By following the methods in [15], the new form is:

$$\Delta\mathbf{y}_{\mathbf{F}}^{(i)}(k) = \zeta_s^{(i)}(k-d)\mathbf{u}_t^{(i)}(k-d) + \mathbf{n}_{\mathbf{F}}^{(i)}(k) \quad . \qquad (3.9)$$

The vectors $\Delta\mathbf{y}_{\mathbf{F}}^{(i)}(k)$ and $\mathbf{u}_t^{(i)}(k-d)$ are known every instant of time, while the scalar $\zeta_s^{(i)}(k)$ is continuously estimated.

The details of the estimation equations are presented in [12]. Further analysis is given in [7] and [15].

Manipulator control for grasping

Manipulator motions are effected by a control law similar to that in the previous sections:

$$\begin{aligned} \mathbf{u}_{con}(k) = {} & -[\hat{\mathbf{J}}^{\mathrm{T}}(k)\mathbf{G}_{\mathbf{M}}\hat{\mathbf{J}}(k) + \mathbf{G}_{\mathbf{I}}]^{-1}\hat{\mathbf{J}}^{\mathrm{T}}(k)\mathbf{G}_{\mathbf{M}} \\ & \{[\mathbf{y}(k) - \mathbf{y}_{des}(k+d)] + \\ & \sum_{m=1}^{d-1}\hat{\mathbf{J}}(k-m)\mathbf{u}_{con}(k-m)\}. \end{aligned}$$

We use this control law during both the object centering and gripper alignment phase, and the object approach and grasping phase. We can also extend the use of the controller to the grasping of moving objects since we consider slowly moving targets. If the higher speed objects exist, the control law can be modified to include the motion of the object as a disturbance term. The values of $\mathbf{y}_{des}(k)$ are held constant during the centering and alignment phase and are time-varying during the approach phase. During approach, several intermediate values of the desired feature point locations are automatically calculated. These intermediate values are used to smoothly guide the gripper to the object and to maintain gripper alignment throughout the approach and grasping phase. Even when the object is in motion, the alignment and centering requirements of the controller cause the manipulator to track the motion, resulting in a system that can grasp objects in spite of their motion.

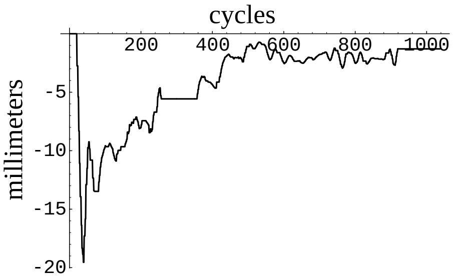

Figure 2: X-axis translation

4. Experimental Results

The proposed work is experimentally verified using the Minnesota Robotic Visual Tracker (MRVT) [4] to automatically select object features, to derive estimates of unknown environmental parameters, and to supply a control vector based upon these estimates to guide the manipulator in the grasping of a moving object.

We assume that the object of interest is a rectangular prism with at least one linear dimension (width or length) that fits into the span of the gripper fingers. We also assume that there are some surface markings that provide suitable fine features for grasping.

We conducted several sets of experiments by varying object's beginning position, orientation, depth, and motion. The first set of experiments was conducted by placing the object approximately 520 mm in depth, 44 mm from the optical axis of the camera, and at a rotation of 14° about the object's Z-axis. The object moves approximately parallel to the Y-axis of the manipulator. During the experiment, the object reverses direction, as shown in Figure 3. The system first aligns the gripper with the minimal linear dimension of the object and forces the optical axis to pass through the centroid of the object. Figure 2 and Figure 3 show the alignment of optical axis and the object centroid during the early potions of the plots. Figure 3 clearly shows the rotation needed to align the gripper with the minimal linear dimension of the object. Figure 5 shows the

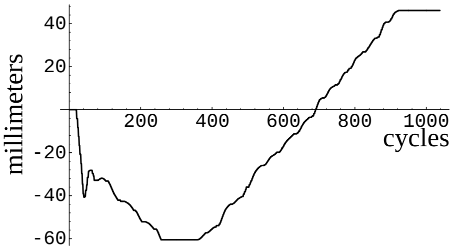

Figure 3: Y-axis translation

reach along the Z-axis (optical axis) and, once the system has identified that the conditions required for grasping are met and the gripper has been closed, shows the withdrawal of the manipulator along the Z-axis near the end of the plot. The time axis is given in cycles that correspond to the cycle time of the robot's controller (28 msec).

During this experiment, object alignment (both gripper alignment and object tracking) is updated as the system is able to measure the position of the features more accurately. Figure 3 reflects the adjustments made during the approach as the object reverses direction.

It should be noted that the system must assume that the coarse feature points are the corners of the object in order to identify the minimal linear dimension and perform the alignment. Also, the size of the black square that supplies the fine points is known, allowing the system to predict the final positions of the fine features in order to allow grasping; however, the system selects these points automatically and their initial positions are unknown and are dependent upon the movements of the manipulator during the coarse guidance and the object's motion.

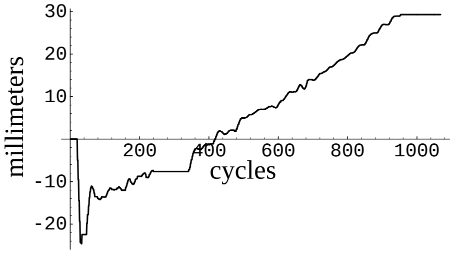

Figure 6: X-axis translation

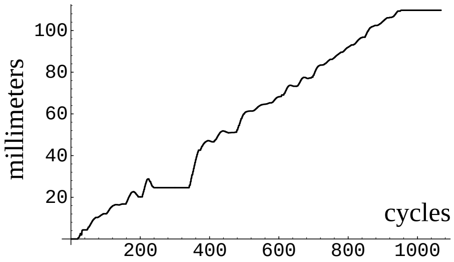

Figure 7: Y-axis translation

The second set of experiments used an object that exhibited a slightly curved path coupled with an object rotation about an axis parallel to, but not colinear with, the optical axis. Figure 6 and Figure 7 show the translational motion of the object. Figure 6 also shows an initial translation to align the optical axis with the object centroid. In this case, the object was initially displaced 20 mm along the X-axis. Figure 8 shows the rotation of the object during the grasping task. The rate of rotation increases over time, resulting in some oscillation during the grasping. Figure 9 shows the approach, grasp, and withdrawal with respect to the Z-axis. In this experiment, the three dimensional motion of the object caused the system to take slightly longer to drive the features to their desired positions; thus, overall time-to-grasp was longer.

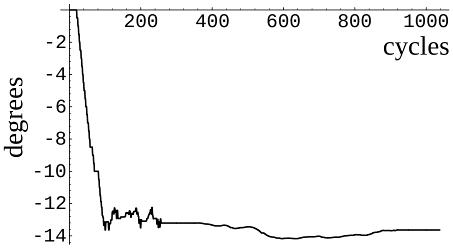

Figure 4: Z-axis rotation

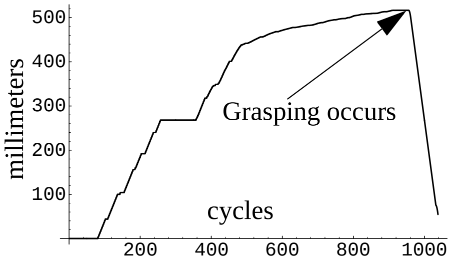

Figure 5: Z-axis translation

In these experiments, the minimal dimension of the object (59 mm) falls within the span of the gripper fingers (72 mm) with only a small tolerance on each side for control error and noise. If the system simply relied upon the gripper and optical axis alignment of the first stage, the accuracy needed to grasp a wide object such as this box would not have been available and the failure rate of the system would have been much higher. The object motion in both of these experiments also demonstrates the benefit of using vision throughout the grasping task. The change in direction of the first experiment and the non-constant rate of rotation in the second experiment pose significant problems for a system that includes a preprogrammed component during grasping. Since our system has no such preprogrammed component, it successfully grasps the object in both cases.

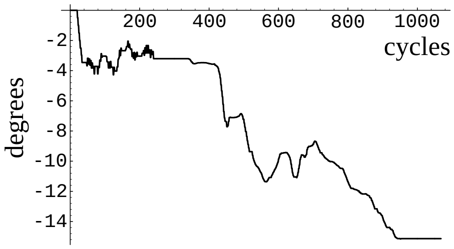

Figure 8: Z-axis rotation

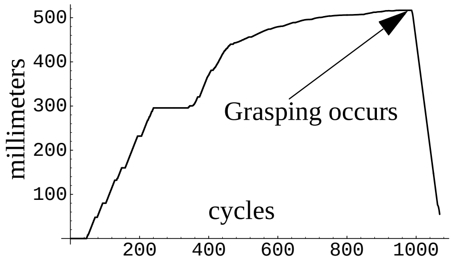

Figure 9: Z-axis translation

5. Conclusions

In this paper, we have presented a method of incorporating visual sensing during basic grasping tasks. This allows a robotic system to achieve a high level of accuracy while grasping objects that are near in size to the gripper opening. The method is based upon the Controlled Active Vision framework [12] and is implemented using the Minnesota Robotic Visual Tracker [4].

The system successfully grasps rectangular prisms regardless of the initial orientation and motion, even though the objects used have only a single graspable dimension that requires extremely tight tolerances to fit within the gripper's fingers. It also does not require a calibrated camera or accurate measurements of other environmental parameters (e.g., focal length, tool transformation, object dimensions, etc.).

The preliminary system uses *a priori* knowledge about the fine feature points and is currently restricted to a single geometric class of objects. These limitations provide the basis for future work and refinements to the system.

6. Acknowledgments

This work has been supported by the Department of Energy (Sandia National Laboratories) through Contracts #AC-3752D and #AL-3021, the National Science Foundation through Contracts #IRI-9410003, #CDA-9222922, and #IRI-9502245, the Center for Transportation Studies through Contract #USDOT/DTRS 93-G-0017-1, the Minnesota Department of Transportation through Contracts #71789-72983-169 and #71789-72447-159, the Army High Performance Computing Center and the Army Research Office through Contract #DAAH04-95-C-0008, and the McKnight Land-Grant Professorship Program at the University of Minnesota.

7. References

[1] P. Allen, A. Timcenko, B. Yoshimi, and P. Michelman, "Automated tracking and grasping of a moving object with a robotic hand-eye system," *IEEE Transactions on Robotics and Automation*, 9(2):152-65, 1993.

[2] P. Allen, B. Yoshimi, and A. Timcenko, "Real-time visual servoing," *Proc. of the IEEE International Conference on Robotics and Automation*, 851-856, April, 1991.

[3] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, 2(3):283-310, 1988.

[4] S. Brandt, C. Smith, and N. Papanikolopoulos, "The Minnesota robotic visual tracker: a flexible testbed for vision-guided robotic research," *Proc. of the IEEE International Conference on Systems, Man, and Cybernetics*, 1363-1368, 1994.

[5] F. Chaumette, P. Rives and B. Espiau, "Positioning of a robot with respect to an object, tracking it and estimating its velocity by visual servoing," *Proc. of the IEEE International Conference on Robotics and Automation*, 2248-2253, 1991.

[6] J. Feddema and C. Lee, "Adaptive image feature prediction and control for visual tracking with a hand-eye coordinated camera," *IEEE Transactions on Systems, Man, and Cybernetics*, 20(5):1172-1183, 1990.

[7] G. Goodwin and K. Sin, *Adaptive filtering, prediction and control*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1984.

[8] N. Houshangi, "Control of a robotic manipulator to grasp a moving target using vision," *Proc. of the IEEE International Conference on Robotics and Automation*, 604-9, 1990.

[9] A. Koivo, "On adaptive vision feedback control of robotic manipulators," *Proc. of the IEEE Conference on Decision and Control*, 2:1883-8, 1991.

[10] F. Lewis, *Optimal control*, John Wiley & Sons, New York, 1986.

[11] P. Maybeck, *Stochastic models, estimation, and control*, Academic Press, London, 1979.

[12] N. Papanikolopoulos, "Controlled active vision," Ph.D. Thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, 1992.

[13] N. Papanikolopoulos and P. Khosla, "Robotic visual servoing around a static target: an example of controlled active vision," *Proc. of the 1992 American Control Conference*, 1489-1494, 1992.

[14] N. Papanikolopoulos, P. Khosla, and T. Kanade, "Adaptive robotic visual tracking," *Proc. of the American Control Conference*, 962-967, 1991.

[15] N. Papanikolopoulos, B. Nelson, and P. Khosla, "Six degree-of-freedom hand/eye visual tracking with uncertain parameters," To appear, *IEEE Transactions on Robotics and Automation*, 1995.

[16] A. Schrott, "Feature-based camera-guided grasping by an eye-in-hand robot," *Proc. of the IEEE International Conference on Robotics and Automation*, 1832-1837, 1992.

[17] C. Smith, S. Brandt, and N. Papanikolopoulos, "Controlled active exploration of uncalibrated environments," *Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition*, 792-795, 1994.

[18] C. Smith and N. Papanikolopoulos, "Computation of shape through controlled active exploration," *Proc. of the IEEE International Conference on Robotics and Automation*, 2516-2521, 1994.

[19] C. Smith and N. Papanikolopoulos, "Grasping of static and moving objects using a vision-based control approach," *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 329-334, 1995.

[20] D. Westmore and W. Wilson, "Direct dynamic control of a robot using an end-point mounted camera and Kalman filter position estimation," *Proc. of the IEEE International Conference on Robotics and Automation*, 2376-2384, 1991.