

# Vision-Guided Robotic Grasping: Issues and Experiments

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

*Many researchers have turned to sensing, and in particular computer vision, to create more flexible robotic systems. Computer vision is often required to provide data for the grasping of a target. Using a vision system for grasping presents several issues with respect to sensing, control, and system configuration. This paper presents some of these issues in concert with the options available to the researcher and the trade-offs to be expected when integrating a vision system with a robotic system for the purpose of grasping objects. The paper includes experimental results from a particular configuration that characterize the type and frequency of errors encountered while performing various vision-guided grasping tasks. These error classes and their frequency of occurrence lend insight into the problems encountered during visual grasping and into the possible solution of these problems.*

## 1 Introduction

In the field of robotics, sensors have long been viewed as providing the solution to many difficult problems involving the interaction of the robot with its environment. Rarely has this promise been met. Typically, sensing fails either from overly ambitious goals for the sensing system or from unanticipated problems relating to the sensor, the robot, and the integration of their respective systems. In this paper, we address the issues related to sensing for robotic tasks by analyzing a standard robotic sensor problem: providing vision-based data for the grasping of a target.

In most grasping applications, the desired effect of incorporating a vision sensor into the task structure is to increase the success rate of the robotic grasp. We assume that the need for a sensor implies that the position of the object with respect to the manipulator is unknown or partially known, otherwise the sensor would be providing redundant information. Certainly there are other reasons for sensing, such as object detection, object recognition, etc.; however, it is the uncertainty of where the object is located that typically dominates the decision to use vision-based robotics.

In this paper, we present an overview of the work related to vision-guided grasping and the design decisions researchers made when addressing sensor system and robot system issues. We then re-examine these issues from

a design viewpoint and present some of the alternatives and trade-offs that are key to these issues. Next, we highlight the design choices we have made while developing a vision-guided grasping system, driven in part by the goal of high success rate grasping. We also present the results of extensive runs of grasping experiments, including a discussion of the problems we encountered and our solutions to those problems. Finally, we summarize our results and present the unresolved issues related to vision-based grasping in general and our system in particular. We also use these issues as the motivation for future work.

## 2 Prior Research

Prior work in the use of visual information specifically for grasping has resulted in only a limited number of efforts. Many of these proposed systems used a static camera and a calibrated coordinate transformation from the camera frame to the manipulator frame. Additionally, these systems typically used open-loop control. In particular, several efforts used vision only to gather position information before performing a blind grasp.

Houshangi [4] developed a system to grasp targets exhibiting planar, translational motion. The system utilized a static camera and pre-computed manipulator poses for the start position and the grasp position. This simplified the issues of calibrated camera-to-manipulator transformations, but limited the flexibility of the system. It was assumed that the target's geometry and pose were known and that the object maintained constant velocity (i.e., no speed nor direction changes). By limiting these factors, the control of the manipulator was simplified, and the control system could accurately predict object position for a given point in time. There was no discussion regarding the use of visual feedback during the grasping motion.

Koivo [6] proposed a control theory approach for grasping but no experimental results were given. This work was similar to Houshangi's and possessed the same trade-offs and issues.

Kimura *et al.* [5] proposed a system capable of catching a free flying ball using a four-degree of freedom manipulator. The target's geometry (a sphere) rendered the object graspable from any approach angle and thus eliminated any issues regarding approach angles and alignment constraints. Additionally, since the object was tossed, the target's trajectory was assumed to be planar, translational motion of known acceleration (specifically, acceleration

due to gravity). This allowed the prediction of the trajectory by the vision system, reducing the system's dependency upon a real-time vision system. The static camera simplified the control algorithms, but raised a calibration issue for precise camera-to-manipulator coordinate system transformation. A simple frame-differencing technique was used to find motion energy and to estimate the object position. This solved many problems that arise when using vision for real-time control, but could only be used with a single moving object.

Schrott [7] proposed a set of actions that an eye-in-hand system should do in order to grasp a static known object. The choice of an eye-in-hand configuration reduced the need for a calibrated camera-to-manipulator coordinate transform. It also closed the control loop by providing visual feedback (robot motion causes camera motion) without requiring the vision system to track the end-effector (required for closed-loop control with a static camera). Issues of real-time control and vision system delays were not addressed and experimental results were not given.

Buttazzo *et al.* [3] used a calibrated static camera to determine when and where an object would cross a line defined by the intersection of the operating plane of the manipulator and the ground plane of the workspace. This information was used with open-loop control to place a basket mounted on the end-effector over the object when the object crossed the predefined line. The choice of this particular configuration required an accurate camera-to-manipulator coordinate transform in addition to the camera calibration required to determine ground plane to workspace transforms. The use of a basket as the end-effector eliminated gripper alignment issues, allowed some calibration error, and tolerated noise in both sensing and manipulator positioning.

Allen *et al.* [1] presented a system that tracks a target moving in an oval path and grasps the target when tracking becomes stable. The visual component of the system depended upon frame differencing to determine the location of the moving target, reducing vision processing time to acceptable levels for real-time control. Blobs were produced using successive frames from two cameras, the centroids of the blobs were calculated, and the geometry of the stereo camera pair was used to determine the 3-D location of the target. This location was used to drive the manipulator in tracking above the moving target. This configuration eliminated problems with moving cameras, but introduced the need for stereo-baseline calibration and camera-to-manipulator coordinate transform calibration. The graspable dimension of the target was assumed to be normal to the tangent of the curvilinear target motion, thus Z-axis rotation was assumed to be dependant upon X- and Y-axis translation. The grasping motion used no sensor feedback. This reduces the accuracy of the grasp, introduces the potential for impact of the gripper with the ground, and eliminates the ability of the system to com-

pensate for errors in calibration, for noise in the sensor, and for changes in target path after the reach is initiated.

### 3 Issues for Vision-Based Grasping

In this section we explore the issues and trade-offs that must be considered when designing a system to perform vision-based grasping of objects. Several of the relevant issues have been introduced in the discussion of the previous work in the area. We will attempt to clarify the issues that need to be addressed, offer some of the trade-offs related to these issues, and detail our own design decisions that were made during the development of our vision-guided grasping system [9][10] based upon the MRVT [2].

#### 3.1 Open-Loop Versus Closed-Loop Control

One of the most basic design issues in any vision-based robotic system is the choice of open- or closed-loop control. Many grasping systems use open-loop control to eliminate the need to either sense the end-effector via the vision system, or to use an eye-in-hand system where manipulator motion introduces camera motion. Often, open-loop control is required since the processing of the vision system is too slow for real-time control or the vision system is used only to determine the object's position and orientation prior to a blind grasp. Closed-loop control requires that the visual data is used as a feedback signal in the manipulator control and requires vision processing with acceptable speed and delay factors for real-time applications. Closed-loop control also allows visual data to compensate for manipulator positioning inaccuracies and (to a limited extent) sensor noise. Typically, open-loop control requires more accurate calibration of the camera-manipulator system while closed-loop control requires more and faster vision hardware.

If closed-loop control is selected, then the selection of a control algorithm and the incorporation of the visual data into the control scheme becomes an issue. Trade-offs between adaptive versus non-adaptive control arise as do alternatives for the objective of the controller and for the measurement of error in the system.

For our system, we chose an adaptive, closed-loop controller to reduce calibration requirements and to provide what we considered a better potential for the subsequent extension of the system to moving objects that exhibit non-constant motion [9][10]. The controller is based upon a repositioning controller that attempts to drive object feature points to selected positions on the image plane (see [10] for a discussion of the control law).

#### 3.2 Blind or Visually-Guided Grasps

Many vision-based robotic systems for grasping attempt to determine the 3-dimensional workspace of the manipulator via vision and then perform a blind grasp of the target object. This decision relaxes real-time constraints since the manipulator is not guided to the object by vision data, but it places a premium on the calibration and accuracy of the vision system and the certainty with which

the camera's position is known relative to the manipulator. Small errors in calculated object positions can cause the system to fail when attempting the blind grasp.

Visual guidance of the grasp offers a system that can more easily recover from the effects of sensor noise and errors in calibration and can more easily be adapted to the grasping of objects exhibiting unknown, non-constant motion. However, it requires that the vision processing period and delays are appropriate for real-time control.

We emphasize visual guidance throughout the grasp since we believe that this offers the best possibility for successful grasps under uncertainty with respect to manipulator positioning, visual data, and calibration. The vision algorithms and control law we use for accomplishing this are given in [9][10].

### 3.3 Monocular, Stereo, or Structured Light

Stereo and structured light systems share the issue of baseline calibration. While stereo systems can be designed that tolerate small errors in calibration, the canonical stereo system requires baseline calibration and solution of the correspondence problem. Likewise, structured light systems typically use a laser stripe and a camera with a calibrated baseline to determine the three-dimensional position of an object. To achieve equivalent performance, the stereo system requires either twice as much vision processing hardware, or hardware that is twice as fast as that required by a structured light system.

A monocular system shares the vision hardware advantage that structured light systems hold over stereo. In addition, there is no baseline that requires calibration. Unfortunately, monocular systems cannot provide the three-dimensional location of an object without employing some method to recover depth (see [8]) or it must assume a calibrated ground plane that is not parallel to the image plane of the camera. Typically, monocular systems attempt to define tasks by relating the motion of features on the image plane to the motion of objects in the workspace. We have selected a monocular approach [9] to grasping that utilizes a repositioning controller to effect changes in the pose of the manipulator with respect to an object to be grasped. This choice was made to reduce hardware requirements, to eliminate the correspondence problem, and to eliminate baseline calibration.

### 3.4 Camera Placement

The choice of a mounting position for the camera can cause drastic changes in the basic system design. Camera positioning usually involves a choice between a statically mounted camera and a robot mounted camera (eye-in-hand configuration). The static mount removes the problems associated with a moving camera, but introduces other issues. For instance, issues regarding camera-to-manipulator coordinate transformation calibration arise. Further, the manipulator may occlude the object in some configurations. Finally, a method for deriving the three-

dimensional position of the object (e.g., ground plane, stereo, structured light) must be chosen.

The robot-mounted camera can eliminate the need for accurate calibration, stereo or structured light systems (in most cases), and assumptions about the workspace (e.g., ground plane assumptions). The robot mount has problems with a moving camera (if the actual grasp is to be vision-guided), a wide range of operating depths dependent upon the manipulator configuration, and a potential for the manipulator's position to cause lighting changes.

We have chosen to use an eye-in-hand configuration for several reasons [9]. Primarily, we want a closed-loop system, but we do not wish to incur the cost of using the vision system to sense the position of the end-effector, as discussed previously. We also wish to avoid expensive and possibly repetitive calibration processes that could affect the reliability of the system. Finally, we hope to avoid occlusion of the object view by part of the manipulator by selecting this configuration.

### 3.5 Object Geometry

The choice of what type of objects to grasp can change the complexity and generality of any grasping system. For example, a spherical object renders certain aspects of alignment constraints irrelevant (i.e., gripper alignment with a graspable dimension of the object). Cylinders allow the object to roll without changing the appearance of the object (if they have a uniform, untextured surface). Both cylinders and prisms require that the system align or pre-shape the gripper to match the dimension of the object that is graspable. Arbitrary objects require object recognition or extensive grasp planning in order to grasp the objects.

We use rectangular prisms with one graspable dimension [9][10], thus requiring gripper alignment without needing a grasp planner or object recognition system (at this point). This choice balances object complexity and system complexity in a way that emphasizes vision-guided grasping, rather than recognition or planning.

## 4 Experimental Results

As presented in previous work, we have implemented and tested our system for visually-guided grasping on both static-object and moving-object grasps. This earlier work gave preliminary results of the system, but it did not categorize failure types and analyze the results of extended numbers of random experiments. We performed this type of analysis on static-object grasps in order to evaluate the system and suggest modifications to improve system performance. When system performance met our criteria for success, we tested the system on a preliminary series of moving-object grasps to determine the success rate.

Three sets of experiments were conducted. The first set exposes a problem in the feature selection and reselection process. The second set demonstrates that the improved version meets our target success rate for static grasping. This set also suggests improvements that might be made to

the system. The third set shows that the system performance for moving grasps achieves promising results during initial experiments.

#### 4.1 Experimental Design

The object’s position with respect to the end-effector was varied by a vector  $(\Delta x, \Delta y, \Delta z)$  and the rotation about the Z-axis (optical axis) was varied by an amount  $\Delta\Theta$ . Each delta was derived using a uniform random number generator and was bounded by a maximum and a minimum chosen for each delta. The deltas were applied to an initial position that corresponded to the object centroid aligned with the optical axis ( $\Delta x = 0$  and  $\Delta y = 0$ ), the graspable dimension aligned with the gripper ( $\Delta\Theta = 0$ ), and with an object depth of 581.7 mm.

We identified three basic failure categories that correspond to prior observed system behavior. The categories are 1) bad feature point selection/reselection; 2) loss of tracking for one or more feature points; and 3) feature point passing out of the view of the camera.

#### 4.2 Initial Experimental Run

Our initial experiment run consisted of 64 random object placement experiments. Each of the 64 experiments had a random variation in all four dimensional variables of interest. Table 1 shows the ranges for each dimensional variable. The results of the 64 runs are shown in Table 2. The results of these experiments uncovered a systemic error in the feature selection/reselection algorithm. When the depth of the object was less than a specific amount during the selection of fine features or when the features were severely distorted during reselection, the algorithm would incorrectly chose a degenerate (untrackable) feature. It should be pointed out that some of the “lost track” failures could also be due to a poor feature selection or reselection. Since overall performance was inadequate and a known problem existed, the problem was corrected and another run of random object placement experiments to determine overall system performance was performed.

Dimension	Minimum	Maximum
$\Delta x$	-38.2 mm	56.8 mm
$\Delta y$	-78.0 mm	77.9 mm
$\Delta z$	-184.7 mm	135.3 mm
$\Delta\Theta$	-28.7 deg	25.0 deg

Table 1: Variable Ranges, Experimental Run 1

Result	Number of instances out of 64
Success	43
Bad selection	16
Lost track	5

Table 2: Outcomes, Experimental Run 1

Result	Number of instances out of 64
Out of view	0

Table 2: Outcomes, Experimental Run 1

#### 4.3 System Performance Experimental Run

A second experimental run of 100 random object placement experiments was conducted. The range of variation of some of the dimensional variables was expanded for the second run since the failure rate of experiments where the confirmed reason was not the systemic error was exceptionally small. Table 3 shows the range of variation of the dimensional variables, while Table 4 shows the failure categories and the frequency of occurrence over the 100 random runs. Performance of the system exceeded the success rate desired (90%) for static grasping.

Dimension	Minimum	Maximum
$\Delta x$	-40.0 mm	60.0 mm
$\Delta y$	-89.2 mm	86.9 mm
$\Delta z$	-149.8 mm	149.7 mm
$\Delta\Theta$	-29.9 deg	29.4 deg

Table 3: Variable Ranges, Experimental Run 2

Result	Number of instances out of 100
Success	96
Bad selection	1
Lost track	5
Out of view	0

Table 4: Outcomes, Experimental Run 2

The data from a typical random object placement experiment is shown in Figure 1 through Figure 4. In these plots, *cycles* refers to Puma controller cycles of 28msec each. The X- and Y-axis rotational dimensions are not relevant in this experiment and are omitted since their magnitudes are both smaller than one-tenth of one degree. The plot for the Z-axis translation (Figure 3) has been

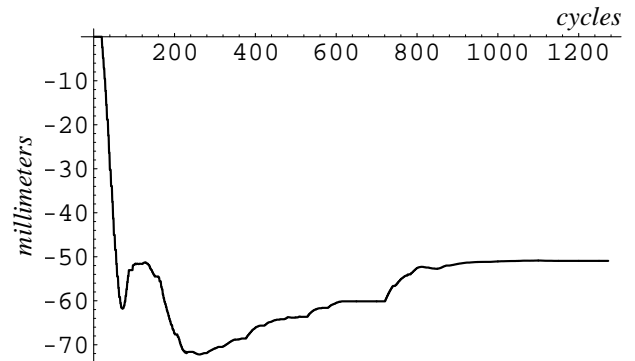


Figure 1: X-Axis Translation

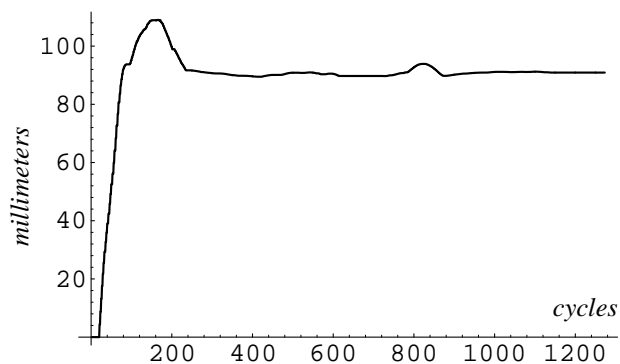


Figure 2: Y-Axis Translation

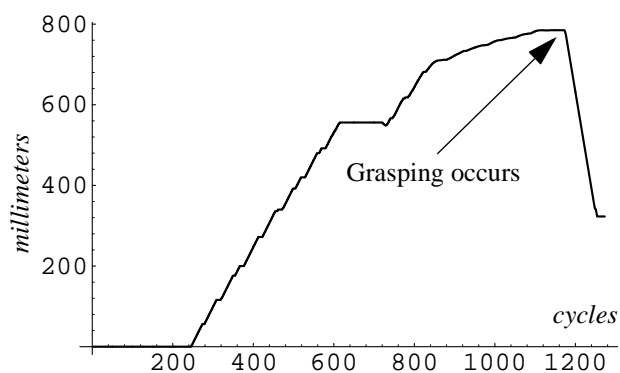


Figure 3: Z-Axis Translation

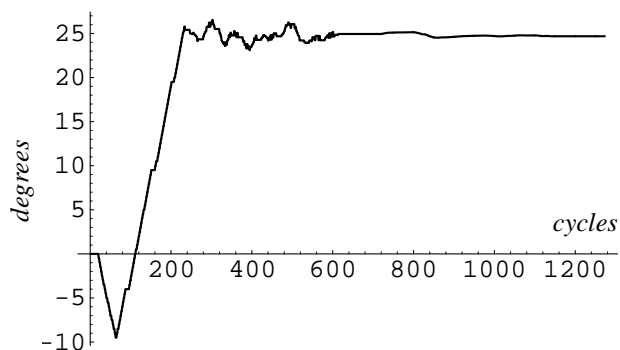


Figure 4: Z-Axis Rotation

annotated to show the moment when the gripper is closed and the manipulator withdraws along the Z-axis.

#### 4.4 Moving Object Experimental Run

Considering the success of the static object grasping after identifying problems and incorporating changes, we applied the revised grasping system to the problem of moving target grasping. A limited run of 10 experiments was conducted to examine the suitability of the changes to the moving grasping problem. The experiments used the same method as the previous static runs, randomizing the initial object placement over the four variables.

Table 5 shows the range of variation of the dimensional variables, while Table 6 shows the failure categories and their frequency of occurrence over the 10 trials. These pre-

liminary results are encouraging, considering that the same control law, without a disturbance term, was used.

Dimension	Minimum	Maximum
$\Delta x$	-40.0 mm	44.6 mm
$\Delta y$	-88.6 mm	78.2 mm
$\Delta z$	-135.9 mm	147.3 mm
$\Delta\Theta$	-26.0 deg	25.8 deg

Table 5: Variable Ranges, Moving Experimental Run

Result	Number of instances out of 10
Success	8
Bad selection	0
Lost track	2
Out of view	0

Table 6: Outcomes, Moving Experimental Run

Figure 5 through Figure 8 show the motion of the arm during the grasping task. Figure 5 and Figure 6 indirectly indicate target motion. Again, the X- and Y-axis rotational dimensions are not relevant in this experiment and are omitted since their magnitudes are both smaller than one-twentieth of one degree. The rate of rotation increases over time, resulting in some oscillation during the grasping. Figure 7 shows the approach, grasp, and withdrawal with respect to the Z-axis.

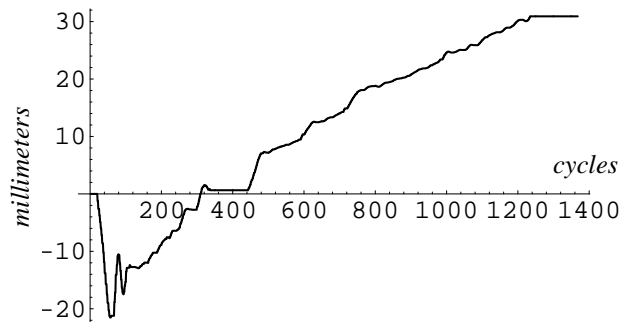


Figure 5: X-Axis Translation

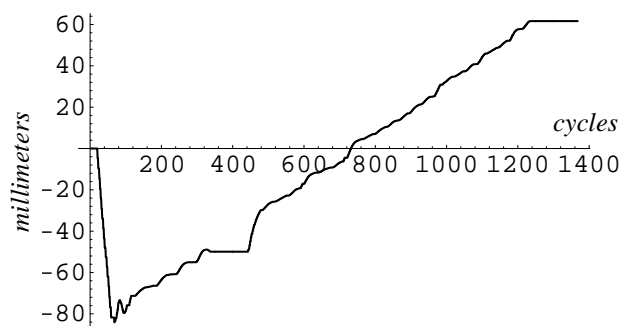


Figure 6: Y-Axis Translation

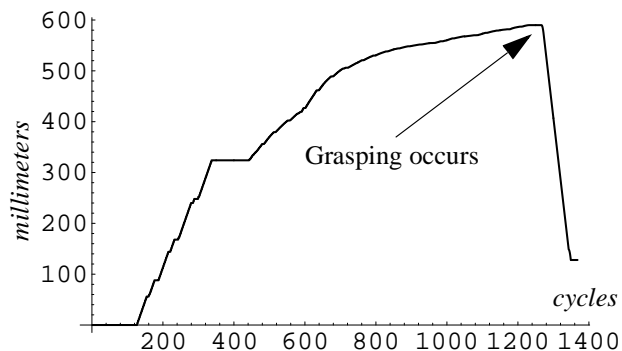


Figure 7: Z-Axis Translation

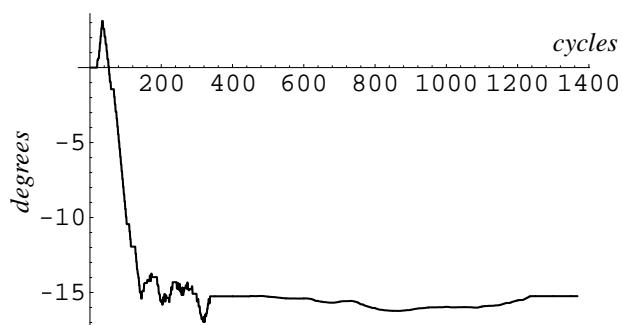


Figure 8: Z-Axis Rotation

## 5 Conclusion

In this paper, we have presented important issues related to the system design using a vision-based approach to robotic grasping. We have framed the issues with the trade-offs that are related to each one of these issues. In addition, we have presented prior research efforts in this area and detailed how the various issues and trade-offs were treated by each of these previous efforts.

Our own approach to these issues has been presented, including the rationale behind the decisions that were made when we were designing our grasping system. We also present the results from three sets of random object placement experiments that emphasize failure analysis in order to improve the system's performance. After the first set of random experiments, we identified and fixed an intermittent, systemic error in our feature selection/reselection scheme. The second run demonstrates the efficacy of our approach and meets our initial requirement of a success rate  $> 90\%$ .

The same system achieved promising results when applied to the grasping of a moving object, with a preliminary run of random object placement experiments showing an 80% success rate.

By analyzing the remaining failures in both static and moving object grasping, we have identified a problem with the camera position with respect to the gripper. We have also observed a systematic loss of tracking just after the selection of the fine features (see [9][10]) that indicates a problem with the gains in the adaptive controller. This

analysis provides the basis for our future work on the grasping 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 and #IRI-9502245, the Army High Performance Computing Research Center under the auspices of the Department of the Army, Army Research Laboratory cooperative agreement number DAAH04-95-2-0003/contract number DAAH04-95-C-0008 (the content of which does not necessarily reflect the position of the policy of the government, and no official endorsement should be inferred), the University of Minnesota Graduate School Doctoral Dissertation Fellowship Program, 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):152-165, 1993.
- [2] S. Brandt, C. Smith, and N. Papanikolopoulos, "The Minnesota Robotic Visual Tracker: A Flexible testbed for vision-guided robotic research," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 1363-1368, 1994.
- [3] G. Buttazzo, B. Allotta, and F. Fanizza, "Mousebuster: a robot system for catching fast moving objects by vision," *Proceedings of the IEEE International Conference on Robotics and Automation*, 3:932-937, 1993.
- [4] N. Houshangi, "Control of a robotic manipulator to grasp a moving target using vision," *Proceedings of the IEEE International Conference on Robotics and Automation*, 604-609, 1990.
- [5] H. Kimura, N. Mukai, and J. Slotine, "Adaptive visual tracking and Gaussian network algorithms for robotic catching," *Winter Annual Meeting of the American Society of Mechanical Engineers*, 43:67-74, 1992.
- [6] A. Koivo, "On adaptive vision feedback control of robotic manipulators," *Proceedings of the IEEE Conference on Decision and Control*, 2:1883-1888, 1991.
- [7] A. Schrott, "Feature-based camera-guided grasping by an eye-in-hand robot," *Proceedings of the IEEE International Conference on Robotics and Automation*, 1832-1837, 1992.
- [8] C. Smith and N. Papanikolopoulos, "Computation of shape through controlled active exploration," *Proceedings of the IEEE International Conference on Robotics and Automation*, 2516-2521, 1994.
- [9] C. Smith and N. Papanikolopoulos, "Grasping of static and moving objects using a vision-based control approach," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 329-334, 1995.
- [10] C. Smith and N. Papanikolopoulos, "Theory and experiments in vision-based grasping," *Proceedings of the 34th IEEE Conference on Decision and Control*, 4053-4058, 1995.