AN INTELLIGENT VISION GUIDED TELEROBOTIC SYSTEM FOR FILE MANIPULATION AND OFFICE AUTOMATION

Kevin G. Stanley, Jonathan Wu
Vancouver Innovation Centre
National Research Council
Vancouver, BC V6T1W5 Canada

William A. Gruver
School of Engineering Science
Simon Fraser University
Burnaby, BC V5A1S6 Canada

Abstract

This research describes a vision guided telerobotic system that enables people with disabilities to perform clerical or office tasks. By adding a light duty robot to the office workspace the operator can manipulate files and perform other work related tasks. To increase the effectiveness of the robot, vision can be used to verify that the robot is correctly positioned. In addition, vision can be be coupled with the telerobotic system to allow the user more intuitive control over the robot. Visual servoing and traditional computed kinematics actions are inappropriate for this application because visual servoing requires an excessive number of iterations, and computed kinematics requires accurate calibration. To counteract these difficulties, and provide user functionality, we have designed a hybrid computed kinematics telerobotic system with an initial coarsely calibrated computed kinematics step followed by a more accurate visual servoing step. We show that there are significant performance benefits from this approach. Finally, we describe how the hybrid system may be utilized in an office environment.

1. INTRODUCTION

Vision guided robotics has a rich research history that dates back to the late seventies and early eighties. While many elements of vision guided robotics have been thoroughly researched, few vision guided robotic systems are used in industrial tasks. Many vision guided robotic systems are too slow, and too sensitive to the environment to be useful. With the rapid rise in computing power and the lower prices of high quality robotic and vision systems, the application of vision guided robotic systems in industrial tasks is becoming more common. There are still barriers and limitations, however, to the utilization of vision guided robotics. In this paper we propose a hybrid approach to vision guided robotics.

An overview of visual servoing is given by Hutchinson, *et al.* [1] who describes the research and fundamentals of geometric feature based visual servoing. Corke [2] has shown that the performance of visual servoing algorithms can be enhanced by incorporating the dynamics of the system in the model. Papanikopolous, *et al.* [3] have used adaptive control techniques to perform visual servoing.

Calibration of the kinematic transforms between the image and the world coordinate system has been examined by many researchers. The calibration-based method has been based on determining the coefficients of the transform between the robot and camera frames. Wang [4] described in detail the relationships between the frames and applied three different methods to approximate the transform, ranging from the case of known target and position, to unknown target and position. Horaud, *et al.* [5] described the effect of a perspective model on the accuracy of the approximation. Wei, *et al.* [6] outlined an approach for computing the transform based on active vision principles. Zhuang, *et al.* [7] described a system where both the robot and camera were calibrated simultaneously. Remy, *et al.* [8] simplified the estimation by employing Euler representations in the transform.

The core problem in vision guided robotics is balancing the requirements for accuracy and speed. Accuracy requires that many samples of the target position be taken throughout the motion of the robot. Speed constraints require that only a small number of iterations are needed before the robot reaches its target. While the cost of computers and robotic systems are rapidly decreasing, it will still be some time before the cost of additional computing power is negligible with respect to the performance gains. Thus, it is necessary to reduce the amount of computation by reducing the number of iterations so that a balance between the speed of execution and the accuracy of positioning is obtained. We propose a system based on a hybrid computed kinematics and visual servoing system.

Our hybrid system is composed of a coarsely calibrated computed kinematics and a gain scheduled PD controller that performs visual servoing. We demonstrate that significant performance gains can be obtained by using a coarse computed kinematics motion prior to starting to servo.

Section 2 contains a description of the components of the hybrid system. Section 3 contains our experimental results, Section 4 contains a description of how the research is being employed in an office environment, and Section 5 discusses conclusions and future research.

0-7803-6583-6/00/$10.00 © 2000 IEEE

2. HYBRID SYSTEM

Problem Statement

The system is designed to perform discrete part manipulation tasks. The workspace is assumed to be locally planar, but the orientation of the plane is unknown. Targets are assumed to be locally planar. A diagram of the system is shown in Fig. 1.

The purpose of the system is to grasp a target at some unknown position $\mathbf{x}_t$ by positioning the end-effector of the robot to a position $\mathbf{r}$ with respect to the target. The position of the robot $\mathbf{r}$ is known from encoder sensing and the inverse kinematics. The target position $\mathbf{x}_t$, must be estimated from the image systems. The x, y, and θ components of the target position $\mathbf{x}_t$ are estimated by measuring the position of the target in $\mathbf{x}_{cr}$ in the end-effector mounted camera frame $\{\mathbf{I}_r\}$. Because targets may have varying heights, we must also estimate the distance along z in the world coordinate frame $\{\mathbf{O}\}$. This measurement is made by examining the image frame of the external camera $\{\mathbf{I}_e\}$. The difference in the y axis with respect to $\{\mathbf{I}_e\}$ between the observed robot position $\mathbf{r}_{ce}$ and the measured target position $\mathbf{x}_{te}$, generates w, the observed difference in world z.

By combining $\mathbf{x}_{cr}$ with w as measured in $\{\mathbf{I}_e\}$ we generate the image space position of the target as

$$\mathbf{x}_{ti} = (u\, v\, w\, \theta)^T \tag{1}$$

By this approach, we can use $\mathbf{x}_{ti}$ to position the robot for grasp planning and grasping. This motion can be generated using either computed kinematics or a visual servoing. In computed kinematics, the position of the target in the world frame $\mathbf{x}_t$ is estimated using a transform $\mathbf{T}$ such that

$$\mathbf{x}_t = \mathbf{T}\mathbf{x}_{ti} \tag{2}$$

Determining the transform T is the key focus of [4]-[8]. Because computed kinematics techniques are so sensitive to the calibration of $\mathbf{T}$, we also utilize visual servoing in which the image position of the target, $\mathbf{x}_{ti}$, is regulated to a desired position $\mathbf{x}_{tid}$, so that the three-dimensional position of the target relative to the robot in the base frame $\mathbf{x}_{rt}$ is known,

$$\mathbf{x}_{rt} = \mathbf{r} - \mathbf{x}_t \tag{3}$$

If

$$\mathbf{x}_{ti} = \mathbf{x}_{tid} \tag{4}$$

then Eq. (3) must be satisfied.

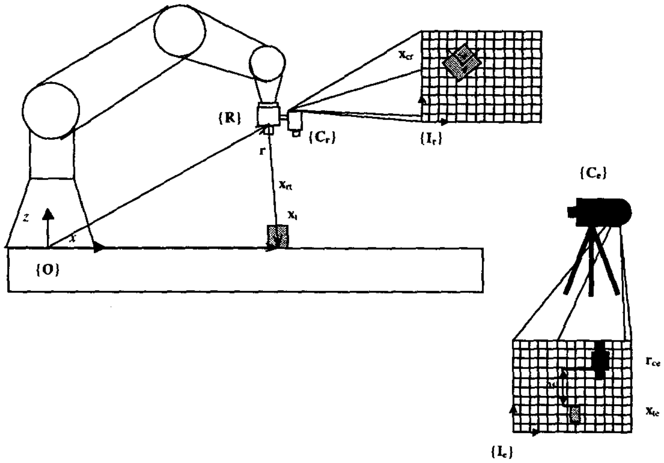

Figure 1: Frame and variable definitions

We have assumed that the parts will be small, discrete, rigid, and stationary. Objects commonly found in industrial assembly or office environments are suitable targets.

Computed kinematics

In computed kinematics vision guided robotic systems, the pose of the target is calculated from the image, and the robot is commanded to move to some point in space with respect to the target. In order to calculate the pose accurately, the kinematic transformation between the camera and the robot must be known with a high degree of precision, and the camera parameters must also be known with a high degree of accuracy. Several adaptive self-tuning schemes have been suggested [4] - [8].

Unfortunately, as shown in Fig. 2, the calibration is no longer accurate enough to generate a reliable position with respect to the target. There are two sources for this error. First, x and y vary inversely with z, so that errors increase as z decreases. Second, the sensitivity to pitch and yaw errors increases as the end-effector moves closer to the target. Instead of a simple linear transform, the motion is governed by the following

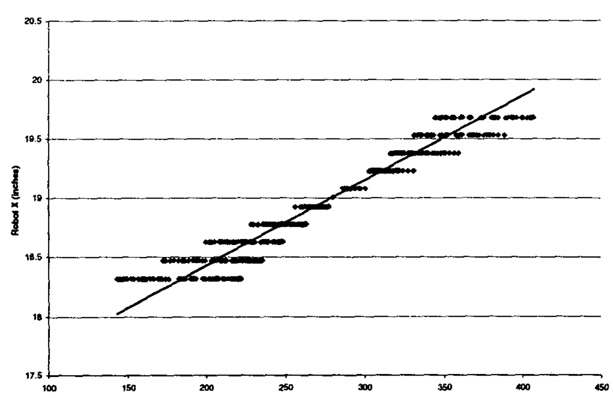

Figure 2: Measured position versus predicted position

$$\mathbf{x}_t = \mathbf{T}\mathbf{x}_{ti} \tag{5}$$

where $\mathbf{T}$ is the transformation matrix between the image and robot frames. The only method for resolving these difficulties when using computed kinematics is to have more accurate calibration. Although six variables of the calibration matrix must be derived, these variables are coupled nonlinearly, thereby increasing the complexity and sensitivity of the calibration procedure [4].

We attempted to compensate the target change in height by applying a pinhole perspective model as described by Horaud, *et al.* [5]:

$$x = c\frac{u}{z} \tag{6}$$

where c is a constant of proportionality. While the addition of the perspective improved the performance it did not completely compensate for the change in perspective because these factors contribute their own errors to the apparent height of the target. From the graph in Fig. 3, however, we noted that the change in depth could be approximated in the local workspace of the robot by a second-degree polynomial

$$w = 6.0\times10^{-5}u^2 - 0.089u + 200.75 \tag{7}$$

which approximates a sinusoid. The maximum variance of this approximation is approximately +/- 5 pixels, which is adequate for our current system needs.

Most computed kinematics systems rely on some form of non-linear programming to perform an approximation of the kinematic mapping. We have taken two approaches to this problem. First, we integrated a neural network approximation $\mathbf{T}$. Then, we included a visual servoing component. Since visual servoing systems are less sensitive to calibration errors, we adopted a secondary visual servoing system to perform the final positioning of the robot.

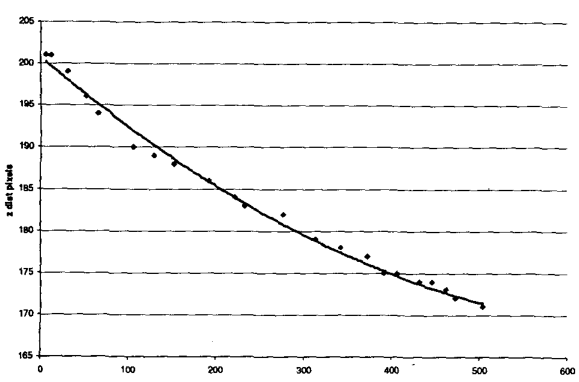

Figure 3: Estimated z calibration

Visual Servoing

In visual servoing a control law is used to position the robot end-effector with respect to a target. Control laws range from simple proportional controllers to complex adaptive schemes [1] – [3]. The control law can be expressed in terms of the image, or in terms of a position in space. In the first case, image-based visual servoing; the control law directly minimizes the error between the current image and the desired image. In the second, Cartesian visual servoing, the pose of the target is calculated and the relative pose between the robot and the target is regulated. The control law may be integrated in the robot controller or it may be implemented as an imaged-based PD control law on an external computer with the original robot controller in the loop. We selected the latter approach for this study.

The most straightforward method for implementation is described by Hutchinson, *et al* [1]. We define the motion of the camera as

$$\dot{\mathbf{C}} = \mathbf{\Omega}\times\mathbf{P} + \mathbf{X} \tag{8}$$

where $\mathbf{C}$ is the speed of the camera, $\mathbf{\Omega}$ is the rotational velocity of the end-effector, $\mathbf{P}$ is the position of the camera with respect to the robot end-effector and $\mathbf{X}$ is the translation velocity of the end-effector. Because we have constrained servoing to 4 DOF, ω_x and ω_y are zero, resulting in

$$\begin{bmatrix} \dot{u} \\ \dot{u} \\ \dot{w} \\ \dot{\theta} \end{bmatrix}_{image} = \begin{bmatrix} \frac{c_x}{z} & 0 & -\frac{u}{z} & -v \\ 0 & \frac{c_y}{z} & -\frac{v}{z} & u \\ 0 & 0 & c_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \omega_z \end{bmatrix} \tag{9}$$

where u is the x coordinate of the image frame, v is the y coordinate of the image frame, and w is the z coordinate of the external image frame. We assume that there is a unity relationship between the observed and actual angle, which should hold true for most cameras. The latter derivation holds for a camera attached off the center of rotation with a constant zoom and a pinhole model. This is an example of a classic Cartesian feature Jacobian for four degrees of freedom. Solving Eq. (9) for the inverse Jacobian, we obtain:

$$\mathbf{J}^{-1} = \begin{bmatrix} \frac{w}{c_x c_z} & 0 & \frac{u}{c_x c_z} & \frac{wv}{c_x c_z} \\ 0 & \frac{w}{c_y c_z} & \frac{v}{c_y c_z} & -\frac{wu}{c_y c_z} \\ 0 & 0 & \frac{1}{c_z} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}$$

The measurement of z in the inverse Jacobian is expressed in terms of w. Because the distance from the manipulator to the top of the target is unknown, we estimate it from measurements made by the external camera.

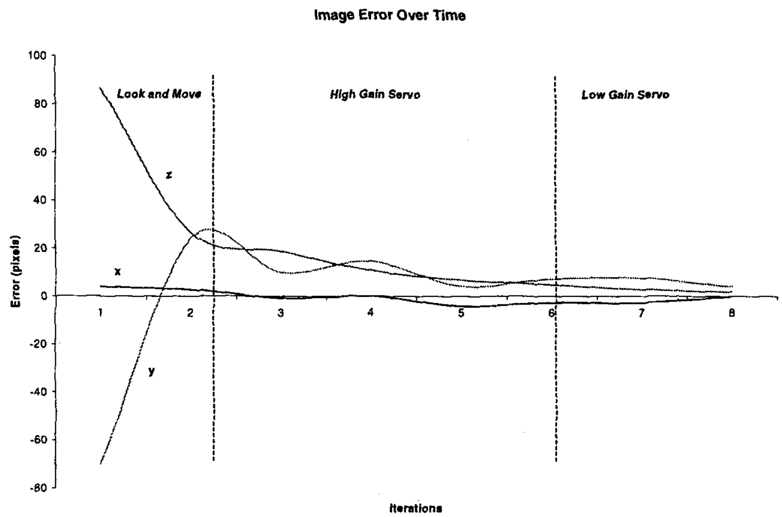

Figure 4: Example system response

Hybrid System Integration

To achieve a rapid response, yet achieve a stable solution, we have employed gain scheduling for our PD controller. The gain scheduling system uses two gain levels, a high gain followed by a low gain to achieve a fast response and accurate final position. Like most high gain PD controllers, our PD controller is characterized by an underdamped response. The purpose of the gain scheduling system is to reduce the oscillations when the error is small.

The lower gain removed the limit cycling behavior, which occurred primarily in x and y, and occasionally caused a coupled effect on θ. Z and θ posed no limit cycling behavior. By decreasing the proportional and derivative gains in x and y by half, we were able to eliminate limit cycling. The ratio between the controller gains in x and y are approximately the same as the ratio between the x and y dimensions of the image. The gains were adjusted so that a similar percent error would evoke a similar response and x and y would converge at approximately the same rate, given the same initial conditions. The high gain system is executed after the initial computed kinematics move. The high gain system continues until the target is within a 15-pixel region of the center of the image. The high gain matrix is then swapped with the low gain matrix. The lower gain matrix has an overdamped response that smoothly drives the robot to its final position. A typical example of the desired system response is shown in Fig. 4.

There is a marked difference in system response for the three stages of motion. The computed kinematics portion is a linear point to point motion, whereas the high gain servo is characterized by a rapid response and oscillatory underdamped behavior. The low gain servo is usually underdamped and moves the robot smoothly to its final position. The three stages of motion are very apparent from the position trajectories.

3. EXPERIMENTAL RESULTS

The hybrid positioning algorithm is compared with a visual servoing control law to evaluate the performance gain of the hybrid system over pure visual servoing using CRS C500 robot. Our first experiment determines the number of iterations and the final accuracy achieved for a circular target for both motion algorithms. The second experiment uses several different objects shown in Fig. 5 to demonstrate the robustness of the system to different targets.

The experimental system was composed of a dual 200 MHz Pentium Pro pc with a Matrox Genesis frame-grabber and digital image processing board on a PCI bus. It was connected to the CRS robot controller running the RAPL-3 operating system via a serial cable. All image processing was accomplished using the Genesis frame grabber. The Genesis board allows multiple threads of processes to be queued via the PCI bus.

In our experiments, the robot was presented with five targets as shown in Fig. 5. The object, a paper cup, appears as a circle in the top projection, and was run eight times for each type of robotic positioning system as a baseline for observing the other targets. The paper cup target was used to test the accuracy of the system. Of the eight target positions, each system was run until an accuracy of ± 5 pixels on each joint was achieved. The final error and number of iterations required are shown in Table 1.

Table 1: Comparison of systems

Algorithm	Final Error	Iterations
Visual Servo	2.98	9.86
Computed kinematics Hybrid	2.73	6.5

We observe that the visual servoing system has approximately the same final error as the hybrid system. However, the computed kinematics hybrid system requires 33% fewer iterations to converge, resulting in a time saving of over a second for this low bandwidth system.

The generalization of the algorithm with respect to the target type was tested with different targets in the workspace. The results are shown in Table 2.

Table 2: Target generalization

	Visual Servoing		**Computed Hybrid**	**kinematics**
	Iterations	Error	Iterations	Error
backing	11	0.79	7	2.48
crushed	12	2.93	12	2.67
elbow	9	3.36	5	3.06
plug	10	2.95	5	3.79
Average	10.5	2.51	7.25	3.00

It is apparent from Table 2 that there is little difference in accuracy (less than one pixel) between the two systems, even with different targets. There is a slight difference between the average values for the cup, and the measured values for the other targets. However, the performance advantage of the hybrid systems over the visual servoing system is significant.

4. OFFICE WORKCELL INTEGRATION

The proposed hybrid computed kinematics system plays an important role in the implementation of a vision-guided office automation workcell. The cell provides a means for manipulating common office equipment. Most of the workcell operation involves the interpretation of voice commands, using commercial voice recognition software, that command the robot to follow set paths. However, the repeatability of the robot is not high enough to reliably handle the required tasks. To compensate, the hybrid system described above is used to determine whether the robot has been homed by checking the image position of a calibration target against a known baseline.

More importantly the computed kinematics system allows the user to perform vision guided teleoperation. To perform actions outside of the standard pre-programmed file manipulations, the user can employ a telerobotic interface to move the robot directly. This interface can be utilized in the following three ways:

1. *Direct robot commands:* The user can command the robot to move to a specific position in the workspace. While this method is accurate it is cumbersome and unintuitive.
2. *Fuzzy direction commands:* The user can command the robot using fuzzy commands associated with a direction such as "move very far left, very fast". While this latter approch is more intuitive than specifying the position, each degree of freedom must be moved independantly.
3. *Vision guided commands*: Using the proposed hybrid computed kinematics system the user can command the robot to grasp a target. In the two-camera example above, the robot positions itself with respect to the robot. If the external camera is occluded the user must provide a height.

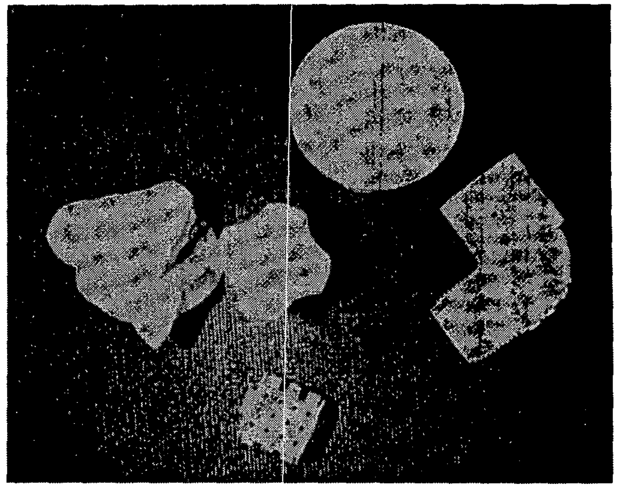

Figure 5: Targets used in Experiments

By allowing the user to select the target, target recognition may be removed from the vision guidance problem. In addition, the user's ability to supply external information, such as the height of the target when the external camera is occluded, provides additional redundancy not possible in fully automated systems.

After reaching the target, the robot can either be provided with grasp points by the user, or determine the grasp points using a planer grasp planning system [9]. Again, by employing the robot in a telerobotic system, the user can improve the workcell performance.

5. CONCLUSIONS

The system described in this research can manipulate a wide range of discrete parts in four degrees of freedom without models of the targets. We demonstrated that an efficient approach to vision guided robotics with limited bandwidth is a hybrid computed kinematics and visual servoing system. Such limited bandwidth applications may be the only applications for visual servoing in industry for the near future because robotic vendors will not provide complete access to their joint level controllers. Furthermore, low cost vision systems will not have the required ability to operate at 1 kHz cycle times needed for servoing. We have also demonstrated that grasp planning can be performed by examining the image without model reconstruction. The technology and hardware cost now allow vision-guided robotics in practical applications. The proposed system has commercial potential for both telerobotics and sorting.

REFERENCES

1. Seth Hutchinson, Gregory Hager, and Peter I. Corke, "A Tutorial on Visual Servo Control", *IEEE Transactions on Robotics and Automation*, Oct 1996, pp. 651-669.
2. Peter I. Corke, Malcolm C. Good, "Dynamic Effects in Visual Closed Loop Systems," *IEEE Transactions on Robotics and Automation*, Oct 1996, pp. 671-683.
3. Nikolaos P. Papanikolopoulos, Pradeep K. Khosla, "Adaptive Robotic Visual Tracking: Theory and Experiments," *IEEE Transactions on Automatic Control,* March 1993, pp. 429-445.
4. Ching-Cheng Wang, "Extrinsic Calibration of a Vision Sensor Mounted on a Robot," *IEEE Transactions on Robotics and Automation,* Vol. 8, No. 2, April 1992, pp. 161-175.
5. Radu Horaud, Fadi Dornaika, Bart Lamiroy, and Stephane Christy, "Object Pose: The Link between Weak Perspective Paraperspective and Full Perspective," *International Journal of Computer Vision* 22 (2), 1997, pp. 173-189.
6. Guo-Qing Wei, Klaus Arbter, and Gerd Hirzinger, "Active Self-Calibration of Robotic Eyes and Hand-Eye Relationships with Model Identification," *IEEE Transactions on Robotics and Automation,* Vol. 14, No. 1, February, 1998, pp.158-166.
7. Hanqi Zhuang, Kuanchih Wang, Zvi S. Roth, "Simultaneous Calibration of a Robot and a Hand-Mounted Camera," *IEEE Transactions on Robotics and Automation,* Vol. 11, No. 5, 1995, pp. 649-660.
8. S. Remy, M. Dhome, J. M. Lavest, N. Daucher, "Hand-Eye Calibration," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1997 pp. 1057-1065.
9. Kevin Stanley, Q.M.J. Wu, Ali Jerbi and William A. Gruver "A Fast Two-Dimensional Image Based Grasp Planner" *Proc. of the IEEE International Conference on Intelligent Robots and Systems,* Kyongju, Korea, Oct. 1999.