MACHINE VISION BASED TELEOPERATION AID

William A. Hoff, Lance B. Gatrell, John R. Spofford
Martin Marietta Astronautics Group
Mail Stop 4372, P.O. Box 179
Denver, CO 80201
Phone: (303)977-4114

ABSTRACT

When teleoperating a robot using video from a remote camera, it is difficult for the operator to gauge depth and orientation from a single view. In addition, there are situations where a camera mounted for viewing by the teleoperator during a teleoperation task may not be able to see the tool tip, or the viewing angle may not be intuitive (requiring extensive training to reduce the risk of incorrect or dangerous moves by the teleoperator). A machine vision based teleoperator aid is presented which uses the operator's camera view to compute an object's pose (position and orientation), and then overlays onto the operator's screen information on the object's current and desired positions. The operator can choose to display orientation and translation information as graphics and/or text. This aid provides easily assimilated depth and relative orientation information to the teleoperator. The camera may be mounted at any known orientation relative to the tool tip. A preliminary experiment with human operators was conducted and showed that task accuracies were significantly greater with than without this aid.

Keywords: Machine Vision, Teleoperation, Telerobotics, Pose Estimation.

1. INTRODUCTION

Telerobotics has the potential to greatly benefit many space applications, by reducing the great cost and hazards associated with manned flight operations. For example, space assembly, maintenance, and inspection tasks can potentially be done remotely using robots instead of using extra-vehicular activity (EVA). Teleoperation is an attractive method of control of such robots due to the availability and maturity of the technology. Unfortunately, using remote camera views degrades the operator's sense of perception as compared to actually having the operator physically on the scene. This paper describes how artificial intelligence (specifically, machine vision) can be used to implement a teleoperator aid that improves the operator's sense of perception.

1.1 The Problem of Perception in Teleoperation

In this paper, we are concerned with the class of teleoperation tasks that involves placing the end-effector of the robot in a certain pose (position and orientation) relative to some other object in the scene. This class of tasks includes most manipulation tasks, since generally an object must be grasped or manipulated in a specific manner, and so the accurate placement of the end-effector with respect to that object is a required precondition. In addition to manipulation requirements, the end-effector must be moved accurately around the workspace to avoid collisions. We are also concerned with the class of tasks in which the identity, geometry, and appearance of the object to be manipulated is well known in advance, but its location is only approximately known. Finally, we are concerned with tasks where the end-effector must be placed relative to the object with an accuracy that is tighter or more stringent than the initial *a priori* knowledge of the location of that object. This situation is common when the task and environment are fairly well specified in advance, but the exact locations of objects are uncertain due to manufacturing tolerances, measurement uncertainties, thermal effects, *etc.*

To provide an illustrative example, many proposed space robotics tasks, such as engaging bolts and mating connectors, are estimated to require positional accuracies of as tight as ±0.125" and ±1° [GSFC87]. On the other hand, the absolute positioning accuracy of the Flight Telerobotic

Servicer (FTS) robot is required to be only ±1", ±3° (this refers to the position of the end effector relative to the robot's stabilizer arm attachment point). In addition to this, there are uncertainties in the robot's docking or attachment mechanism, and uncertainties in the position of the object in the workspace. These can potentially add several inches and degrees to the total end effector-to-object uncertainty.

The net result is that the motion of the robot cannot be pre-programmed in advance, but that some sort of sensor feedback must be used to correct for positioning errors. In the case of teleoperation, the sensor feedback to the operator usually consists of remote camera video and force reflection. Force reflection is only useful when the end effector has already made contact with the object, and may be used in some cases to correct for very small positioning errors. However, to move the end effector from a potentially distant position to close proximity of the object while avoiding obstacles (at which point a "guarded" move can be performed), visual feedback is necessary.

For monoscopic vision, using a single wrist mounted camera (or possibly a head camera in addition to a wrist camera), the operator may have difficulty in perceiving the three dimensional relative position and orientation of objects in the scene. Absolute distances and orientations are even more difficult to judge. In addition to these problems, in some cases the cameras are not able to get a good view of critical locations such as the tool tip, due to occlusions. Extensive operator training can alleviate these problems, but this training must be very specific to a particular task and workspace. Any changes to either the object or the background can mislead the operator and cause errors.

1.2 Approaches to Solving the Perception Problem

There have been a large number of studies performed over the years on the effects of the characteristics of video displays on the ability of the operator to perform manipulative tasks. A single video display usually does not provide good task perspective. Generally, studies have found that stereoscopic vision is superior to monoscopic for typical manipulation tasks because it provides a better sense of depth perception [Pepp83]. Two separate, preferably orthogonal, camera views can also be used to give more perspective. However, this approach requires the operator to look at two displays, and worksite camera locations may be limited. Also, two camera views of equal resolutions can require twice the communications bandwidth on a single camera, which is an important consideration in remote space operations.

Both perspective and stereoscopic visual cues have been shown to improve manual tracking and manipulation ability. A perspective cue that provides information not directly indicated by the task image appeared to improve performance the most in simulation studies [Kim87a],[Kim87b]. A 2D display of a 3D tracking task tends to increase errors of translation into the image and rotations about axes in the image plane [Mas89]. Graphics superimposed on a video view has been previously shown to assist operators in grasping objects with a manipulator [Kim89]. That work used an *a priori* modelled location of the camera and manipulator base to draw a graphic aid based on the current manipulator pose.

1.3 The Machine Vision Approach

In this paper, we propose an alternative solution to the perception problem that is based on the use of artificial intelligence. Specifically, we use machine vision to automatically recognize and locate the object in the camera views, and then display the pose of the object as an overlay on the operator's live video display. The operator can thus teleoperate the robot guided by the computed pose information, instead of or in addition to the video image. We have implemented this system and have measured its benefit to teleoperators in a preliminary experiment.

The remainder of this paper is organized as follows. Section 2 describes previous approaches to teleoperator aids and discusses the advantages of our approach over existing techniques. Section 3 describes our teleoperator aid that we have implemented and evaluated at Martin Marietta. Section 4 summarizes the machine vision technology that is used to derive the object's pose. Section 5 summarizes the remainder of our laboratory facilities, including the robot manipulators, controllers, and operator workstation. Section 6 describes the experiment that was conducted to measure the effect of the teleoperator

aid on operator performance. Section 7 gives conclusions.

2. PREVIOUS TELEOPERATOR AIDS

Without a teleoperator aid, the operator must rely on his visual memory of where the end effector should be in the images to gauge his accuracy. In our task, described in Section 6, there were no stadiametric background marks with which to estimate the end effector position. Without a teleoperator aid, we found that the operator was accurate to approximately five millimeters and about two degrees rotation about any one axis. If the operator knew what his pose errors were, he could reduce them to the accuracy limits of the manipulator and the sensor. A sensor which can inform the operator of his pose errors gives the operator the ability to reduce those errors. With such a sensor, one would expect the operator's performance to be more accurate and repeatable since the operator has more accurate feedback information. Furthermore, one would expect that an inexperienced operator using the sensor could achieve better accuracy and repeatability than an experienced operator working from memory alone — at a fraction of the cost for training.

The most basic example of teleoperator aids is the printed transparent overlays used by astronauts operating the RMS arm to grapple an RMS target, and as a backup mode for final rendezvous of the Space Shuttle orbiter with other spacecraft. Each object that is to be used with the transparency aid has its own set of transparencies — typically at least one for full camera zoom, and one for no camera zoom. Tick marks are placed on the transparency so that the operator can easily determine the approximate depth to the object by how many tick marks are filled in the image by the object. The roll of non-symmetric objects can also be determined if radially converging lines are placed on the object. For the case of the RMS target overlays, the RMS is positioned correctly at an RMS target when the target matches the pattern printed on the overlay.

In related work, Bejczy [Bejc80] has described a teleoperator aid that uses four proximity sensors on the end effector of the Remote Manipulator System (RMS) to compute pitch, yaw, and depth information and display it to the operator. The advantages of our system over this work is that we use video sensors, thus achieving a much greater range of operation (depths up to 72 cm versus 15 cm, and yaw angles of [-25°..40°] versus ±15°), we compute all six degrees of freedom instead of only three, and we have a display that is integrated with the video instead of separate.

Another related work of interest was reported by Hartley and Pulliam [Hart88]. Seven display aids for use by pilots of Remotely Piloted Space Vehicles (RPSV) were presented. These display aids were developed during the course of simulations of RPSV tasks at Martin Marietta's Space Operations Simulation laboratory, where a moving base carriage robot was used to simulate RPSV's. Three of the remote pilot aids were reticles that were overlaid on the operator's camera view that graphically gave the operator feedback on the vehicle's pose and trajectory relative to the target. The remaining aids were basically displayed patterns, such as the RMS docking target, which would match the actual target when the RPSV achieved the goal pose — a dynamic version of the transparencies used by astronauts controlling the RMS. Data for these pose and trajectory aids was obtained by reading the moving base carriage pose information. In a real task, this information would have to be sensed by sensors.

One difference in the work of Bejczy compared to that of Hartley is that Bejczy's sensor display was on specialized boxes, whereas Hartley's displays were overlaid on the principal camera view screen. The advantage of on screen display is that the operator does not have to change his view to see the sensor output. Graphic display devices that allow display of live video and overlaid graphics simultaneously are typically higher resolution devices than standard NTSC monitors, and consequently, alignment aids can be more precisely overlaid in the image, further helping to reduce teleoperator errors.

The principal element of our approach is the machine vision based sensor that processes the operator's wrist camera view to compute the transformation (pose errors) between the current end effector position and the goal position. The pose errors are then overlaid on the wrist camera portion of the operator's high resolution monitor as graphics and text.

Some of the advantages of this system are as follows: No additional sensor is required on the manipulator since it uses the operator's video images. All six degrees of freedom of the pose are presented to the operator, with the data display integrated into the operator's screen so that the operator does not have to change his view to see the data. The displays are dynamic in that they are computer generated for the appropriate object as indicated by the operator — without the need for carrying physical transparencies as currently used by the RMS operators. Finally, the computer is able to produce data that is more accurate than the human operator could produce from the same camera view.

3. THE MACHINE VISION-BASED TELEOPERATOR AID

3.1 Coordinate Systems

Figure 3.1-1 illustrates the primary coordinate systems, or frames, that are involved in the computations of our teleoperator aid. In this Section, we use the naming and notation conventions of Craig [Crai89]. The station frame, $\{S\}$, is fixed in the world and is attached to the object that is to be manipulated. The tool frame, $\{T\}$, is attached to the tool or end effector of the robot arm. The desired position of the tool is given by the goal frame, $\{G\}$. Specifically, when the tool is in the desired position, the tool frame coincides with $\{G\}$. The camera frame, $\{C\}$, is attached to a camera that is rigidly mounted on the wrist of the robot arm. When the robot arm is not in the desired goal position, let the tool frame be represented by $\{T'\}$ and the corresponding camera frame by $\{C'\}$, as shown in Figure 3.1-1. The information needed by the teleoperator is how to transform the current tool frame $\{T'\}$ into the desired tool frame $\{T\}$.

We use the following notation conventions: ${}^A_BT$ represents a homogeneous transformation from frame B to frame A, expressed as a 4x4 matrix. ${}^A_BR$ represents the rotation portion of such a transformation (a 3x3 matrix), and ${}^AP_{BORG}$ represents the translational portion (a 3x1 matrix), which is also the location of the origin of the $\{B\}$ frame in the coordinate system of $\{A\}$.

The machine vision system can provide the location of the object with respect to the camera, *i.e.*, ${}^C_ST$ and ${}^{C'}_ST$. To set up our system, we initially moved the robot arm to the goal position and recorded ${}^C_ST$. During operation, we then moved the arm away and tried to determine the transformation necessary to transform the current tool frame into the goal frame ${}^T_{T'}T$, based on an observation of ${}^{C'}_ST$. If we knew the transformation from the tool frame to the camera frame, ${}^C_TT$, we could find the desired transformation as follows:

$$ {}^T_{T'}T = {}^T_CT\ {}^C_ST\ {}^S_{C'}T\ {}^{C'}_{T'}T $$

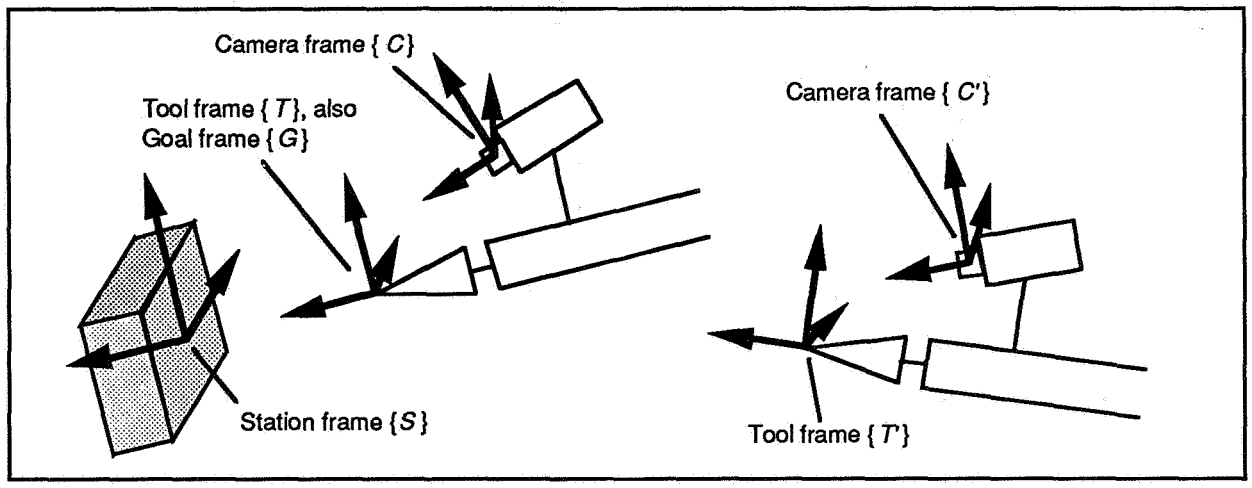

Figure 3.1-1. The primary frames, or coordinate systems, involved in the machine vision based teleoperation aid.

However, in the teleoperator aid implementation described in this paper, we did not know ${}^{C}_{T}T$, although in principle this can be measured. Nevertheless, we were still able to compute useful information about the transformation from $\{T'\}$ to $\{T\}$, because of a special relationship between our station frame $\{S\}$ and our goal frame $\{G\}$. The special relationship is that the axes of these two frames were parallel; *i.e.*, there was only a translation between them. The effect of this was that we were able to compute ${}^{T}_{T'}R$, but not ${}^{T}P_{T'ORG}$. Instead of computing the true translation between the current tool frame $\{T'\}$ and the final tool frame $\{T\}$, the translation that we computed had an additional component due to rotation. However, when there was no rotation, the translation that we computed was correct. This was actually not confusing when we used the teleoperator aid. If we first rotated the end effector to zero the orientation errors, then we could simply translate the end effector according to the displayed pose. Appendix 1 provides a detailed explanation of this. In future work, we plan to measure the camera-to-tool transform so that we can directly compute and display the correct $\{T'\}$ to $\{T\}$ transform.

3.2 Display of Coordinates

Our teleoperation aid displayed the computed pose error as both text and graphics overlaid on the live video. Figure 3.2-1 shows a screen photograph of the text and graphics overlaid on a video image of our truss connector and panel. The lower portion of the screen is occupied by the text. Pose error is reported as the translation in centimeters along the X, Y, and Z axes; and the rotation in degrees about the X, Y, and Z axes (also called pitch, yaw, and roll).

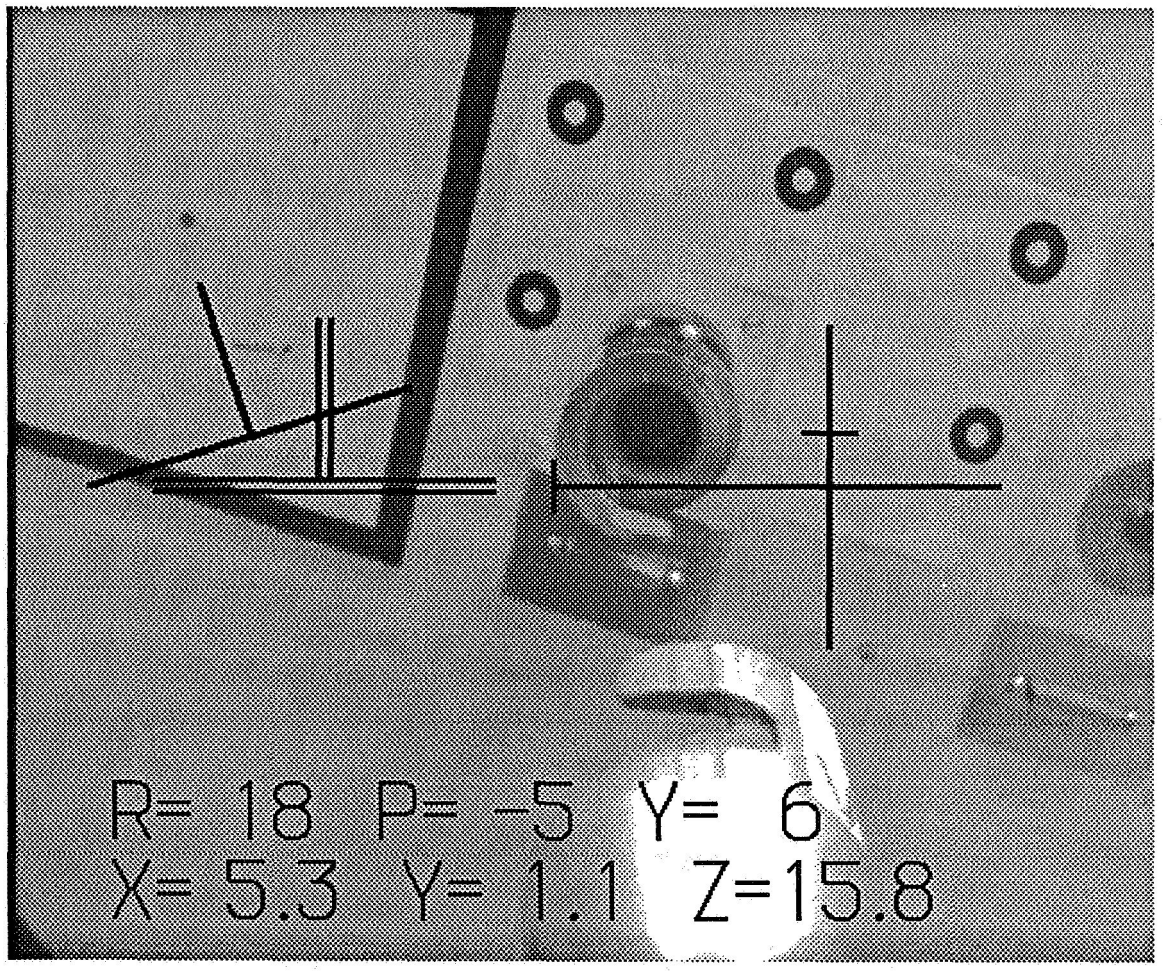

Figure 3.2-1. The teleoperator aid overlaid on live video.

Orientation is also displayed using a reticle, an upside-down "T", in the left center of the screen. Yaw errors are indicated by horizontal translations of the "T", and pitch errors by vertical translations. Roll errors are indicated by a rotation of the "T". By looking at the orientation reticle in Figure 3.2-1, the operator can tell that to get to the goal orientation, he must roll the arm to the right, pitch it down, and yaw to the left. When all errors are zero, the "T" fits exactly inside a hollow fixed reference "T", as shown in Figure 3.2-2.

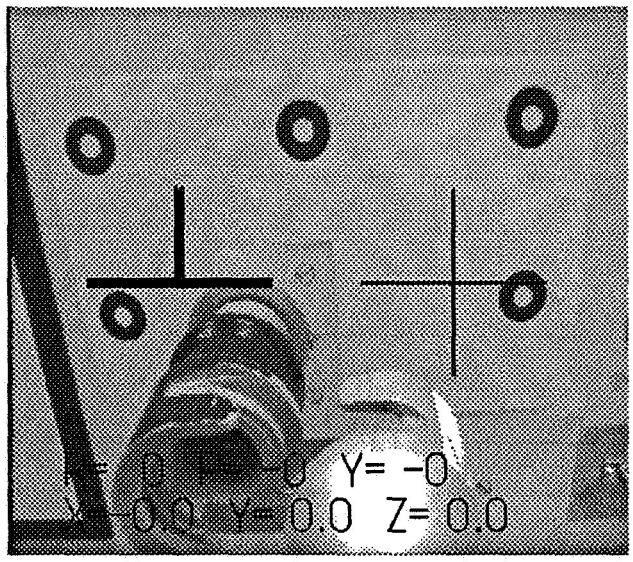

Figure 3.2-2. The display when pose error is zero.

Translation is displayed using a pair of error bars in the right center of the screen, one horizontal and one vertical. Each bar is terminated by a crosspiece at right angles to the bar. Horizontal errors are indicated by a lengthening of the horizontal bar, and vertical errors are indicated by a lengthening of the vertical bar. When the depth error changed sign, the fixed reference crosshairs changed color.

4. MACHINE VISION SYSTEM

The goal of our machine vision system was to identify and estimate the pose of an object of interest in the scene. Although significant progress has been made in the field of machine vision, no system exists at present which can identify large numbers of different objects against multiple backgrounds at video update rates. One alternative is to place visual targets, which can be recognized at video rates, on the objects. Visual targets which have been used to simplify the object recognition process are summarized by Gatrell, *et al.* [Gatr91].

4.1 Image Features

The Concentric Contrasting Circle (CCC) image feature, developed at Martin Marietta and reported in [Gatr91, Skla90], is used in this work for the feature that the machine vision system is looking for. A CCC is formed by placing a black ring on a white background, and is found by comparing the centroids of black regions to the centroids of white regions — those black and white centroids which are equal are CCCs. This image feature is invariant to changes in translation, scale, and roll, and is only slightly affected by changes in pitch and yaw, and can be extracted from the image rapidly with low cost image processing hardware. The centroid of a circular shape is the most precisely locatable image feature [Bose90].

4.2 Object and Target Model

Four CCC's are placed in a flat rectangular pattern on the object to be recognized by our vision system. A fifth CCC is placed on a side of the rectangle to remove the roll ambiguity. This five point target is also described in more detail in [Skla90]. Our basic object recognition process has been reduced to the simple steps of finding five Concentric Contrasting Circles which form a five point Target. We have found this to be very robust and fast. In designing the five point target for a particular object, care must be taken to ensure that all five CCCs will be visible from the expected viewing positions. The target for the truss connector measured 16.5 cm x 6.3 cm; the diameter of the CCCs was 2 cm.

4.3 Pose Estimation

After features of an object have been extracted from an image and their correspondence between image features and object features has been established, the pose of the object relative to the camera, or ${}^{S}_{C}T$, can be computed by many techniques, such as [Chan89, Kris90]. We currently use the simple and fast Hung-Yeh-Harwood pose estimation method [Hung85]. The inputs to the

pose algorithm are the centers of the four corner CCCs, the target model, and a camera model. The pose algorithm essentially finds the transformation which yields the best agreement between the measured image features and their predicted locations based on the target and camera models.

4.4 Camera Calibration

Accurate pose estimation requires an accurate camera model, including the focal length f used above. Most pose estimation techniques, including the Hung-Yeh-Harwood method, assume a pin hole camera model with no lens distortion. Real cameras are not pin hole cameras, and real lenses have noticeable distortion--especially radial lens distortion. The characteristics of a lens, camera, and digitizer configuration are determined by camera calibration. We use the Tsai camera calibration technique [Tsai87] to compute the focal length and radial lens distortion, as discussed in [Skla90]. A picture of the block used to calibrate our camera is shown in Figure 4.4-1. Once again, the image features are the centers of the circles.

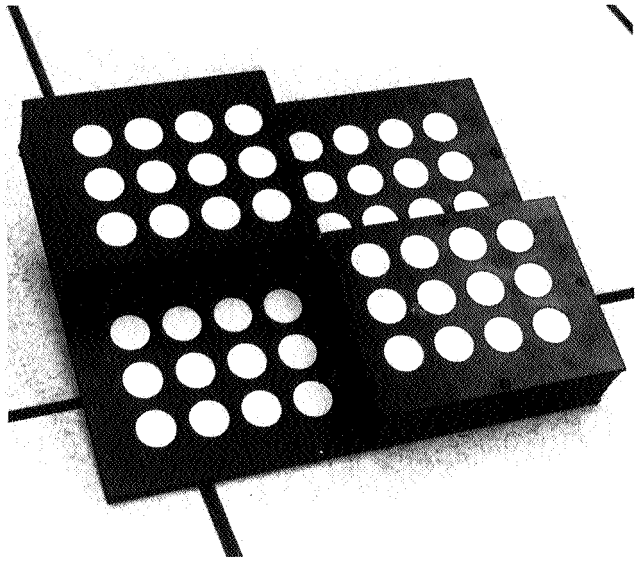

Figure 4.4-1. Block used to calibrate the camera.

During use of our operator aid, the precomputed radial distortion coefficients are used to "undistort" image feature locations on the image plane so that they fit a pin hole camera model. The undistorted image locations are given to the pose estimation routine described above. It should be noted that the Tsai calibration technique also computes the pose of targets relative to the camera, but is not used for estimating the pose of our five point target due to an insufficient number of points for this technique.

4.5 Equipment Used

In our current configuration the wrist camera was a high resolution, black and white, Pulnix TM-840 camera with an 8 mm wide angle auto-iris lens. The shoulder camera was a Panasonic AG-450 S-VHS color camera. Our digitizer and image processing board was an Androx ICS-400XM9. This DSP chip-based board computed the histogram and thresholded the image. The thresholded image was processed on the host machine, a Solbourne 4/501 running at 20 MIPS (Sun SPARC compatible), that computed the connected regions, extracted the CCCs, found the five point target, and computed the pose. The graphics overlays were drawn by the ICS-400XM9 board.

4.6 Vision System Capabilities

The vision system can be in one of three states: 1) not tracking any object, in which case the video is passed through to the operator's monitor, 2) acquiring the object — the last location of the object in the image is unknown, and therefore, the entire image is searched for the object, and 3) tracking the object — the object was found in the last view that was processed, and therefore, each of the target features is searched in a small area centered at the last known location. The acquire step takes about 0.8 seconds once a view with the target visible is acquired. The delay between when a target is visible to the camera and the acquire has finished could take up to 1.6 seconds (up to 0.8 seconds to finish searching the last occluded target view, and 0.8 seconds to process the visible target view). The track step takes about 0.1 to 0.2 seconds per cycle (5-10 Hz.), depending on the size of the CCC target features in the image. As each target feature increases in size, the portion of the image that has to be processed also increases, thus reducing the throughput rate.

A button-input panel on the Solbourne workstation screen allows the user to partially customize the operator aid display and to select from possible options. Aid display features which the op-

erator can choose to have displayed are: 1) colored cross hairs on the CCC centroids, 2) wireframe overlay of the model, 3) orientation reticle, 4) orientation text, 5) translation reticle, and 6) translation text. The operator has the option of setting the current camera to object pose as the goal pose, and the option of reading in a goal pose from a file or writing the current goal pose out to a file, and the option of setting the goal pose to be the object's coordinate frame. Furthermore, the operator can adjust the image contrast with two sliding bars. At present, these capabilities are only selectable on the vision processor workstation screen; our goal is to integrate all of theses features on the high resolution operator console.

With our 8 mm wide angle lens, the target is found at depth ranges of 18 cm to 72 cm. The limiting factor in how close the camera can be to the target is the width of the target — at closer distances the target does not fit in the image. The target can be pitched from -32° to 77°, and can be yawed from -25° to 40°. The pitch and yaw limits are not symmetric because the truss connector extends in front of the target, and as the target pitches and yaws, the connector may occlude some of the CCCs.

5. LABORATORY FACILITIES

This section provides an overview of the facilities in the Combined Robotic Technologies Laboratory (CRTL) where this study was conducted. This laboratory contains a testbed for research in the areas of robotics, teleoperation, applied controls, computer vision, motion planning, situation assessment, and human factors. The laboratory processing architecture allows autonomous and manual control of robotic systems. A more complete description of the laboratory testbed can be found in [Mor90].

5.1 Processing Architecture

Figure 5.1-1 shows the functional architecture for the laboratory testbed. The architecture is based on the NASREM [Alb89] hierarchical model and equivalent NASREM levels are indicated. We have partitioned the system horizontally into three categories; Control and Automation, Situation Assessment and Operator Interfaces. Control and Automation subsystems provide control functions to testbed components or process sensor data used during the control function. Operator Interface subsystems include units which accept operator input and generate commands for Control and Automation units. It also includes units which display information to the operator or record information for analysis. The Situation Assessment subsystem provides the reasoning and model update functions for Control and Automation units which are not used during teleoperation. The study described in this paper used the control functions and operator interfaces at the Prim and Servo levels. A developmental version of the Vision sensing system at the Prim level was used for the operator aid function.

5.2 Robot Manipulators

The CRTL contains two 6 DOF Cincinnati Milacron T^3-726 robot manipulators and one T^3-746 manipulator. All three robots are commercial robots whose control systems have been replaced with custom controllers. They are configured as a dual-arm system with a dynamic task positioning system. Six-axis force/torque sensors are mounted between each manipulator's wrist and end effector. Video cameras are also mounted to the wrist of each manipulator. The servocontrollers for the manipulators receive Cartesian position commands from either the inter-arm coordinator or hand controllers. The commanded position is modified by an impedance control loop, also known as active compliance, based on sensed forces and torques. One of the T^3-726 manipulators was used with a fixed task panel in this study.

5.3 Operator Console

The CRTL has two generations of teleoperation control stations—a three-bay console and a two-bay console. The three-bay console has a center stereo display and two side displays with touch-screen overlays. The center stereo monitor can display live video from a stereo pair of cameras, overlaid with stereo graphics. The operator displays are generated by Silicon Graphics IRIS workstations with video underlays from RGB Spectrum video windowing systems. The resulting displays have 1024 by 1280 pixel resolution and can display video in a variety of resolutions, including a stereo aspect ratio.

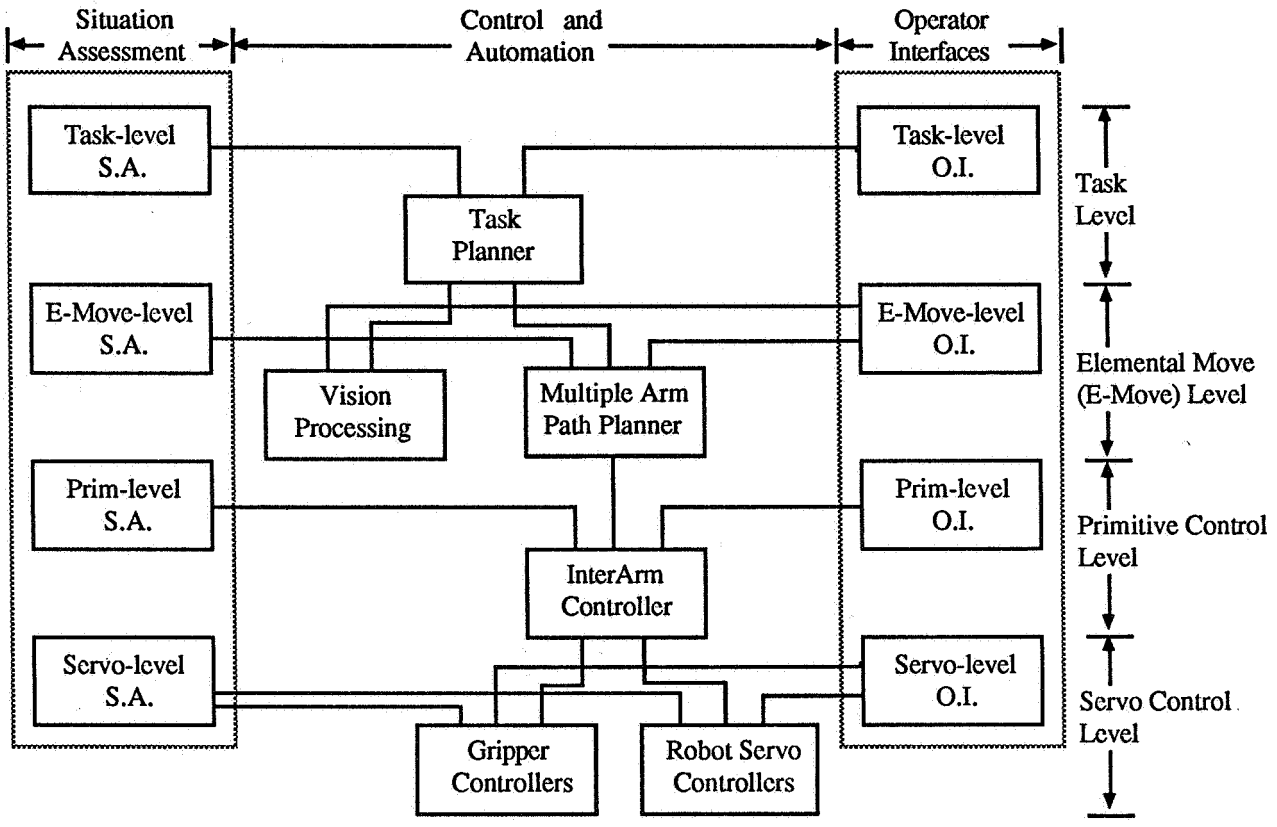

Figure 5.1-1. Functional Architecture for the Tele-Autonomous Testbed

The CRTL control stations are configured to use a wide variety of 6 DOF hand controllers. The Martin Marietta Compact ball was used in this study as the Servo level input device. The Compact ball hand controller is a Cartesian mechanism, with three translational joints and three rotational joints that position and orient the grip, respectively. The mechanical separation of these motions provides natural decoupling between Cartesian axes. A manipulator activation switch on each hand controller enabled manipulator motion when depressed and provided position mode indexing capability.

6. EXPERIMENT

We performed an experiment with human subjects to determine the effectiveness of the operator aid. This was a pilot study, intended to identify areas for more comprehensive future experimentation.

6.1 Experimental Design

Due to time limitations, a single experimental parameter and two subjects were used in this experiment. No comparisons between subjects or conclusions about general populations can be made with this few subjects. However, the experiment was designed so that each subject served as their own control, enabling conclusions about the effectiveness of the aid for each individual subject. The experimental parameter of interest was the presence of the operator aid on the video display, with two levels: present and not. Eight random starting locations were used to form a sample of the subject's performance, thus each subject performed a total of 16 runs. Both subjects were right-handed male engineers with previous experience operating the manipulator and hand controller. Neither had previous experience with the operator aid.

6.2 Task Description

The task selected for this experiment was an unobstructed free-space motion from unknown initial locations to a specified goal location. The task was defined as the final free-space move in a truss node assembly sequence, moving the manipulator-held connector half to a specified position relative to the mating connector half, with no orientation error. Motion in all 6 DOF was required. When the operator aid was displayed, the goal location was indicated by zero position and rotation error. Without the operator aid, only image visual cues were available.

The operator's feedback was from two camera views: the wrist camera was attached to the end effector above and behind it, tilted down at an angle of 18 degrees, while the second stationary camera was placed about 1.2 meters behind and 1.1 meters to the right of the panel mounted truss connector half.

Figure 6.2-1 shows the task panel, the two connector halves, and the camera mounted on the manipulator wrist. The subjects were presented with two video views of the task, one from the fixed color camera and one from the monochrome wrist camera. The manipulator was controlled in tool (end effector) frame, with the hand controller reference frame aligned with the connector seen in the wrist camera view. The tool command frame was not aligned with the wrist camera view, which was pitched down. The hand controller commands were filtered with a cutoff frequency of 4 Hz. Manipulator motion was commanded in position mode, with scaling factors of 0.6 and 0.2 for translations and rotations respectively.

Figure 6.2-2 shows a display similar to that presented to the subjects; the subjects did not see the graphical buttons around the screen periphery. The fixed camera view was full-screen (1:2 pixel mapping) and the wrist camera was quarter-screen (1:1 mapping), located in the upper right corner. Neither the camera views nor the displays could be modified by the operators during testing. The distance and orientation change from each of the eight random starting locations to the goal location are listed in Table 6.2-1.

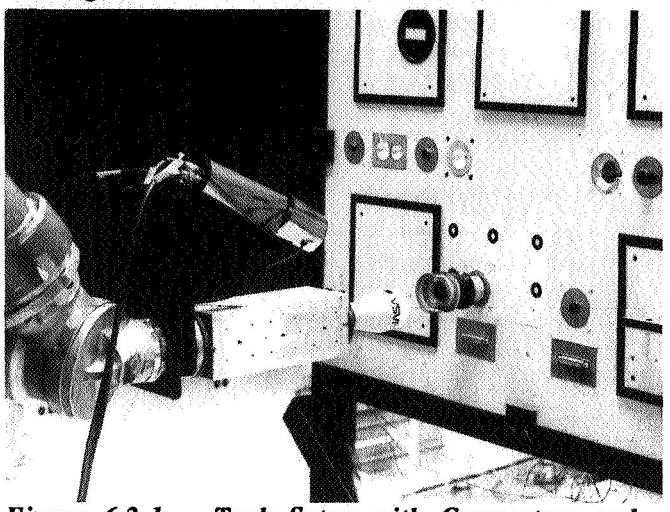

Figure 6.2-1 Task Setup with Connectors and Wrist Camera

Table 6.2-1 Displacement and Orientation from Start Locations to Goal

Start Location	Displacement (cm)	Orientation (deg)
1	19.50	4.46
2	20.82	20.41
3	31.02	20.45
4	26.90	17.72
5	22.61	15.68
6	16.34	15.08
7	21.34	20.81
8	15.55	15.84

6.3 Test Procedure

Each subject performed the experiment as a training session followed by the 16 data collection trials. The whole activity took about 90 minutes per subject. Table 6.3-1 lists the experimental conditions, which were randomly chosen for each subject.

Training was provided to the subjects to familiarize them with the task and display. The subjects were shown the goal position and a subset of the trial starting locations. The subjects were then allowed to operate the manipulator with the hand controller for a self-paced session. They could turn the operator aid on and off between and during training runs.

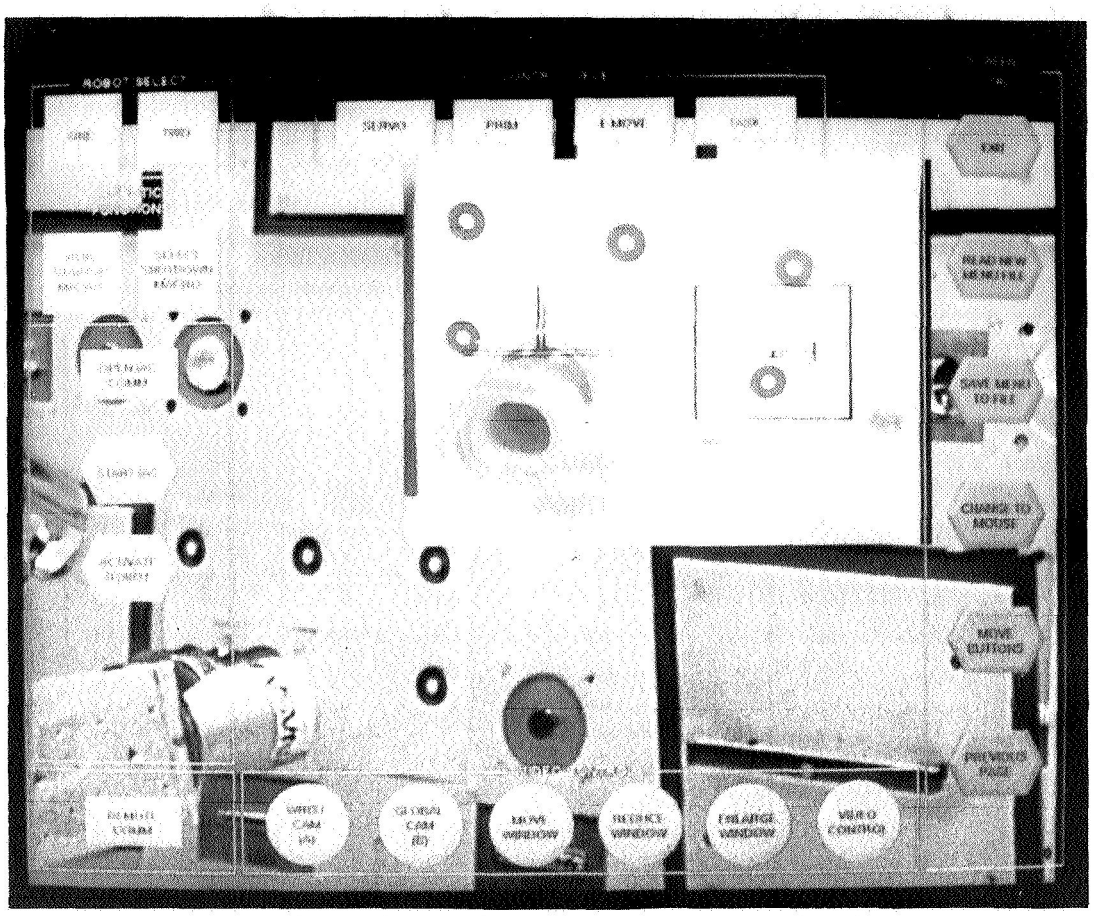

Figure 6.2-2. Screen Display of Fixed and Wrist Cameras

The subjects were able to practice from arbitrary start locations, but not the locations used in the actual experiment. The training sessions lasted until the subjects were comfortable with the system, typically about a half hour. Accuracy was specified as being more important than time in the completion of the task, with magnitudes of 0.2 cm and 1 deg being acceptable.

The trial runs were conducted by using the autonomous control system to move the manipulator to the appropriate start location, and then switching control to the subject in teleoperation mode. The subject was timed using two stopwatches from the start signal until the subject's indication of task completion. The position and orientation errors were calculated from the average of the manipulator's final pose and the vision system's final pose. These generally agreed to within 0.02 cm and 0.2 deg. The subject was not given an indication of performance at the end of each trial, except for the operator aid display when used.

6.4 Analysis Method

The method used here to test for the effect of the operator aid is to test for the possibility that the actual mean for the subject could be the same with and without the aid. This test can be done by comparing the ranges of each performance measure with and without the aid, at a specified See [Hic73] for a more detailed discussion of experimental design and analysis methods.

In the following analysis, the variable x refers to any of the three performance measures: the task time, the position error, or the orientation error. The assumption of a normal distribution for the performance measures was made. Since the sample size is moderately small, a two-tail confidence level of 99% was chosen:

$$P\left[\bar{x}-c<\mu<\bar{x}+c\right] = 0.99$$

This states that the actual mean value μ for each subject is 99% likely to be within c of the sample mean $\bar{x}$, where

$$c = t_{0.995}\frac{s_X}{\sqrt{n}}$$

and $\bar{X}$ and s_X are the mean and sample deviation of the n trials by each subject for each aid condition. $t_{0.995}$ is the value of the *t* distribution with a 99% two-tail confidence level. For a sample with $n = 8$ there are 7 statistical degrees of freedom, and $t_{0.995} = 3.499$.

Table 6.3-1 Experimental Conditions

subject: trial	1 start loc	aid	2 start loc	aid
1	6	YES	1	YES
2	4	YES	2	YES
3	5	no	7	no
4	8	YES	3	no
5	6	no	2	no
6	1	YES	1	no
7	2	YES	6	YES
8	3	YES	6	no
9	7	no	4	no
10	7	YES	5	no
11	2	no	3	YES
12	1	no	8	no
13	4	no	7	YES
14	8	no	4	YES
15	5	YES	8	YES
16	3	no	5	YES

Treating the two aid conditions as separate population samples for each subject, a test can be made to see if they could be samples from the same population. If the 99% confidence ranges of the two samples do not overlap, then we can be confident that the actual means for both populations cannot be the same, indicating that the effect of the operator aid on the performance measure is statistically significant for the specific subject.

6.5 Results

Table 6.5-1 lists the sample mean, sample deviation, and high and low confidence limits of the three performance measures for the two subjects.

The following series of figures shows the performance of the two subjects. Each figure shows the sample mean and high and low 99% confidence values for a performance measure with and without the operator aid.

Table 6.5-1 Experimental Data Summary

		Subject 1 w/ aid	1 w/o aid	Subject 2 w/ aid	2 w/o aid
TASK	$\bar{X}$	47.1	86.1	82.4	63.1
TIME	s_X	13.4	31.4	20.8	32.8
(sec)	$\bar{X}+c$	63.6	124.9	108.2	103.7
	$\bar{X}-c$	30.5	47.2	56.7	22.5
POS	$\bar{X}$	0.20	0.68	0.16	1.18
ERR	s_X	0.08	0.28	0.04	0.56
(cm)	$\bar{X}+c$	0.30	1.03	0.21	1.88
	$\bar{X}-c$	0.10	0.33	0.11	0.48
ANG	$\bar{X}$	0.34	1.80	0.70	3.12
ERR	s_X	0.24	0.74	0.37	0.76
(deg)	$\bar{X}+c$	0.64	2.72	1.17	4.06
	$\bar{X}-c$	0.05	0.88	0.24	2.18

Figures 6.5-1 and 6.5-2 show the task time measure. The average for subject 1 was faster with the aid than without, while the average for subject 2 was somewhat slower. This result will be discussed later. For both subjects, the confidence ranges overlap and no conclusion can be drawn regarding the benefit of the aid. In both cases, the variability of the task time decreased with the aid.

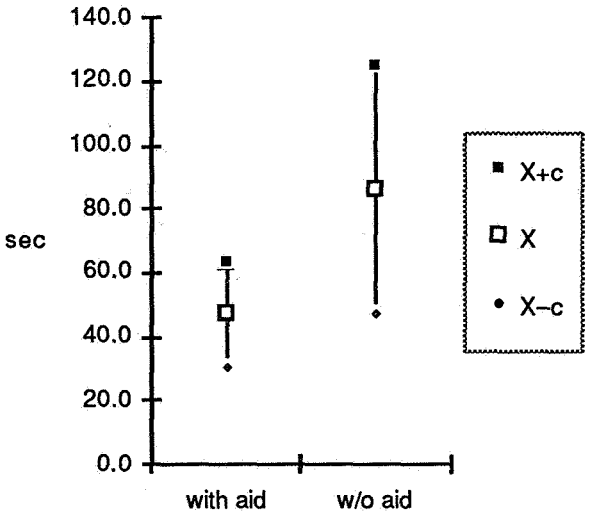

Fig. 6.5-1 Task Time for Subject 1

Figures 6.5-3 and 6.5-4 show the position error measure. Both subjects exhibited less error with the aid, and less variation. The confidences do not overlap, indicating a significant difference related to the presence of the operator aid.

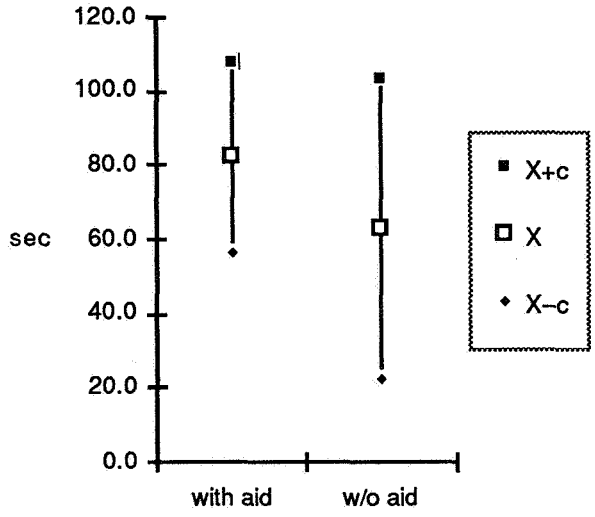

Fig. 6.5-2 Task Time for Subject 2

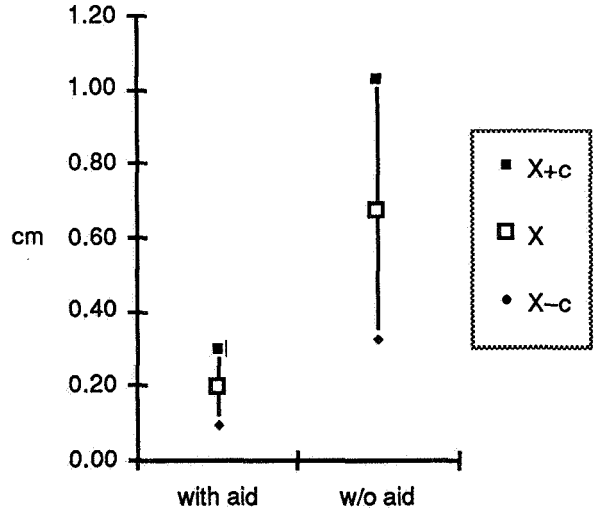

Fig. 6.5-3 Final Position Error for Subject 1

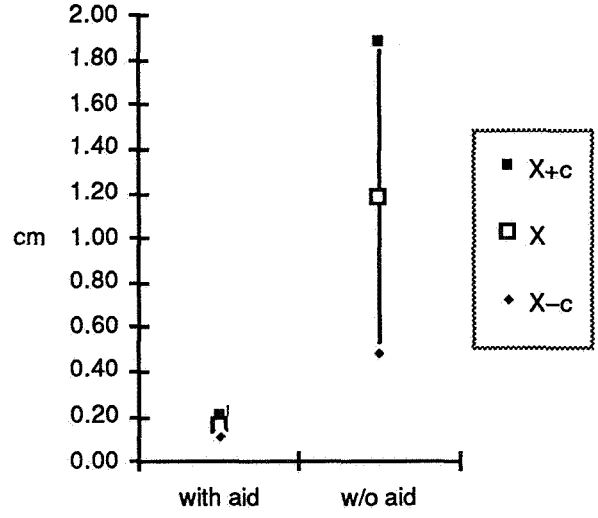

Fig. 6.5-4 Final Position Error for Subject 2

Figures 6.5-5 and 6.5-6 show the orientation error measure. As with the position error, both subjects exhibited less error and less variation with the aid. The confidence ranges do not overlap for either subject, again indicating a significant difference related to the presence of the operator aid.

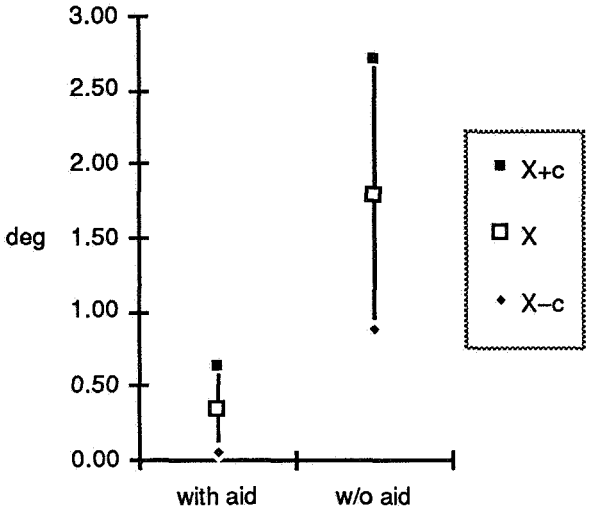

Figure 6.5-5. Final Orientation Error for Subject 1

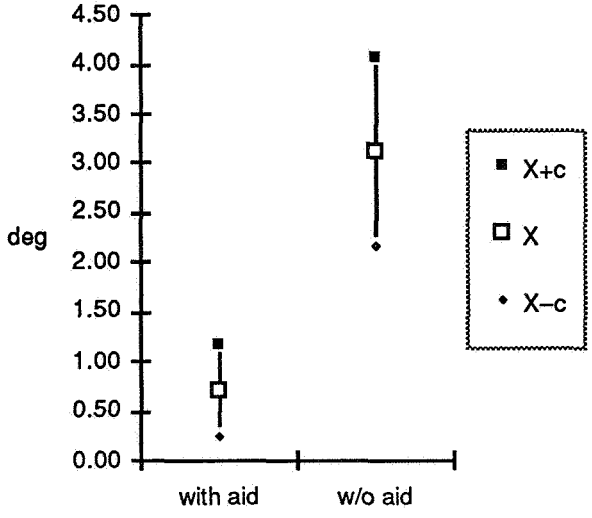

Figure 6.5-6. Final Orientation Error for Subject 2

6.6 Discussion

The subjects made use of geometric features in the images to position and align the connector, they did not just use the image of the connector halves. The accuracy of the final positions would probably have been worse if the vision target had not been visible in the wrist camera view.

The movement style of subject 1 was smoother than that of subject 2, which apparently had an effect on the effectiveness of the operator aid. The aid display dropped out for approximately a second whenever the vision system lost track of the targets. This track loss occurred whenever one target reached the screen edge, or when the image motion between frames was large. The relatively fast motions of subject 2 caused such track losses in seven of the eight trials. This degraded the time performance of the task due to the momentary freezing of the wrist camera view in the current implementation. This may explain why subject 2 took longer to perform the task with the aid present. Subject 1 on the other hand, only experienced one track loss in one trial. It is possible that if the camera view had remained live, with only the aid freezing during a track loss, the average time for subject 2 would have been lower.

In general the accuracy was much better with the aid present, with less definite results for the task time. Since only two subjects were used, no conclusions about the effectiveness of the aid for a general population can be drawn. Based on the results of this experiment however, further study to support more general conclusions appears merited.

7. CONCLUSIONS

In this paper we have described a machine vision based teleoperation aid that improves the operator's sense of perception in performing remote robotics tasks. We have implemented the system and in a preliminary experiment with human operators, have found a significant improvement in their positional accuracy when the aid is used. The aid should be of great benefit to tasks where high accuracy is required, but where there are insufficient camera views, reference markings, etc, to help the operator. The vision based aid also has additional advantages of a large range of operation (in our setup, depths up to 72 cm and yaw angles from [-25°.. 40°]), it is non-contact, and it is integrated with the operator's normal live camera view. Although relatively low cost processing hardware is used, the machine vision system achieves a fairly high update rate of between 5-10 Hz.

This preliminary work has shown the viability of the vision based teleoperator aid and has indicated that there is a great payoff in its use. Future work will follow three directions: (1) improvements in the machine vision system, (2) improvements in the operator's display system, and (3) performance of more extensive experiments.

Improvements to the machine vision system will increase its accuracy, robustness, and flexibility. Specifically, the system currently makes use of a single (wrist) camera view, and locates specially designed optical targets placed on the object. Future enhancements will allow the use of more than one camera view to improve accuracy, and allow a larger set of visual features to be used, which will increase its flexibility and robustness.

Planned improvements to the operator interface include: (1) overlay of the graphical and numerical aids using the IRIS workstation, which will provide live video during track loss and allow positioning of the aids on the entire screen; (2) touchscreen interaction with the vision system to designate target sets of interest in a multi-object view; and (3) stereo presentation of the aids overlaid on stereo video display.

Finally, more comprehensive experimentation will be performed to determine the benefits of different presentations of the vision-derived information, in conjunction with more complex tasks. Communications bandwidth limitations call for a comparison of the relative performance of two orthogonal views, a stereo view, and a single view with aiding.

ACKNOWLEDGEMENTS

The authors would like to thank the test subjects for participating in the experiments. This re-

search was supported by Martin Marietta IR&D projects D-11R and D-75D.

REFERENCES

[Bose90] Bose, Chinmoy B., and Israel Amir, "Design of Fiducials for Accurate Registration Using Machine Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 12, December 1990, pp. 1196-1200.

[Bejc80] Bejczy, A., J. Brown, and J. Lewis, "Evaluation of "Smart" Sensor Displays for Multidimensional Precision Control of Space Shuttle Remote Manipulator," *Proc. 16th Annual Control Conf., Manual Control*, 1980, pp. 607-627.

[Chan89] Chandra, T., and M. A. Abidi, "A New All-Geometric Pose Estimation Algorithm Using a Single Perspective View," *SPIE Vol. 1192 Intelligent Robots and Computer Vision VIII: Algorithms and Techniques (1989)*, pp. 318-329.

[Crai89] Craig, J., *Introduction to Robotics, Mechanics and Control*, Second Edition, Addison-Wesley Publishing Co., Reading, Massachusetts, 1989.

[Gatr91] Gatrell, Lance B., William A. Hoff, Cheryl Sklair, and Michael Magee, "Robust Image Features: Concentric Contrasting Circles and Their Image Extraction," submitted to *Computer Vision and Pattern Recognition*, June 1991.

[GSFC87] Goddard Space Flight Center, "Robotic Assessment Test Sets," GSFC Report No. SS-GSFC-0029, March 1987.

[Hart88] Hartley, Craig S., and Robert Pulliam, "Use of Heads-up Displays, Speech Recognition, and Speech Synthesis in Controlling a Remotely Piloted Space Vehicle," *IEEE Aerospace and Electronic Systems Magazine*, Volume 3, No. 7, July 1988, pp. 18-26.

[Hir89] Hirzinger, G., Heindl, J., and Landzettel, K., "Predicitve and Knowledge-Based Telerobotic Control Concepts," in *Proceedings of 1989 IEEE International Conference on Robotics and Automation*, pp. 1768-1777, May 1989.

[Hung85] Hung, Y., P. Yeh, and D. Harwood, "Passive Ranging to Known Planar Point Sets," *IEEE International Conference on Robotics and Automation*, March 25-28, 1985, pp. 80-85.

[Kim87a] Kim, W.S. et al., "Quantitative Evaluation of Perspective and Stereoscopic Displays in Three-Axis Manual Tracking Tasks," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-17, No. 1, pp. 61-72, January 1987.

[Kim87b] Kim, W.S., Tendick, F., and Stark, L.W., "Visual Enhancements in Pick-and-Place Tasks: Human Operators Controlling a Simulated Cylinderical Manipulator," *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 5, pp. 418-425, October 1987.

[Kim89] Kim, W.S., and Stark, L.W., "Cooperative Control of Visual Displays for Telemanipulation," in *Proceedings of 1989 IEEE International Conference on Robotics and Automation*, pp. 1327-1332, May 1989.

[Kris90] Krishnan, Radha, H. J. Sommer III, and Peter D. Spidaliere, "Monocular Pose of a Rigid Body Using Point Landmarks," submitted to *Computer Vision, Graphics, and Image Processing*, 1990.

[Mas89] Massimino, M.J., Sheridan, T.B>, and Roseborough, J.B., "One Handed Tracking in Six Degrees of Freedom," in *Proceedings of 1989 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 498-503, November 1989.

[Pepp83] Pepper, R., R. Cole, E. Spain, and J. Sigurdson, "Research Issues Involved in Applying Stereoscopic Television to Remotely Operated Vehicles," *Proc. of the Int'l SPIE*, Vol. 402, Geneva, Switzerland, April 1983, 170-174.

[Skla90] Sklair, Cheryl, Lance Gatrell, William Hoff, and Michael Magee, "Optical Target Location Using Machine Vision in Space Robotics Tasks," *Proceedings of SPIE Symposium on Advances in Intelligent Systems*, November 1990, in press.

[Tsai87] Tsai, R., "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 4, pp. 323-344, August 1987.

APPENDIX 1

In our setup, there was no rotation between the station frame and the tool frame in the goal position; i.e., ${}^{T}_{S}T$ was just a translation and ${}^{T}_{S}R = I$. We initially recorded the camera to object transformation when the arm was in the goal position. We then multiplied the current camera-to-object transformation by the inverse of the recorded transformation to yield the pose error. The meaning of this transformation is as follows:

$$\left[{}^{C}_{S}T\right]^{-1} {}^{C'}_{S}T = \left[{}^{C}_{T}T\ {}^{T}_{S}T\right]^{-1} {}^{C'}_{T'}T\ {}^{T'}_{S}T$$
$$= {}^{S}_{T}T\ {}^{T}_{C}T\ {}^{C'}_{T'}T\ {}^{T'}_{S}T$$

The camera is rigidly mounted on the wrist, so the above equation reduces to:

$$\left[{}^{C}_{S}T\right]^{-1} {}^{C'}_{S}T = {}^{S}_{T}T\ {}^{T'}_{S}T$$

We now rewrite the above with the 4x4 homogeneous transformation matrices:

$${}^{S}_{T}T\ {}^{T'}_{S}T = \left(\begin{array}{c|c} {}^{S}_{T}R & {}^{S}P_{TORG} \\ \hline 0 & 1 \end{array}\right)\left(\begin{array}{c|c} {}^{T'}_{S}R & {}^{T'}P_{SORG} \\ \hline 0 & 1 \end{array}\right)$$

$${}^{S}_{T}T\ {}^{T'}_{S}T = \left(\begin{array}{c|c} {}^{T'}_{T}R & -{}^{T'}_{T}R\ {}^{S}P_{T'ORG} + {}^{S}P_{TORG} \\ \hline 0 & 1 \end{array}\right)$$

The correct translational component should be $-{}^{S}P_{T'ORG} + {}^{S}P_{TORG}$. Our translational component has the additional factor of the rotation, ${}^{T'}_{T}R$.