A VIRTUAL REALITY BASED POINT-AND-DIRECT ROBOTIC SYSTEM WITH INSTRUMENTED GLOVE

David J. Cannon, Geb Thomas, Collin Wang, T. Kesavadas
Computer Integrated Manufacturing Laboratory
Department of Industrial and Management Systems Engineering
The Pennsylvania State University
University Park, PA 16802, USA

A system was developed to feature the flexible capabilities of robots by directing them with gesture phrases such as "put that there." This natural human-computer interface to robots uses virtual tools that are graphic representations of robot end-effectors to designate pick, place and other points of interest in live video views of the robot's workspace. Guided by an instrumented glove worn on the operators left hand, these virtual tools "fly" in live video. An operator is able to specify object positions and destinations quickly without resorting to cumbersome telemanipulation or teach pendant control of an actual robot having multiple links. The robot provides trajectory planning between points so the human only needs to point and give directives. Tasks have been developed for hazardous material handling and for rapid development of production systems in flexible manufacturing.* This paper reports our new capability and describes a brief pilot test designed to assess one aspect of using an Instrumented Glove in our Supervisory Control Interface.

Significance:
This virtual reality based point-and-direct (VR-PAD) approach to robotics enhances rapid development of robotic applications in hazardous environments and quickens set-up, on-line error recovery and process modification in flexible manufacturing. The interface frees the operator from details that have previously hindered robot implementation and allows humans and machines to do what each does best. The use of an instrumented glove for our human-machine interface is supported by a speed-of-specification pilot test for a three degree of freedom targeting task.

Keywords:

(Received: 15 April 1994 Accepted: 16 May 1994)

1. INTRODUCTION

The Virtual Reality based Point-And-Direct (VR-PAD) Program, initiated with the Stanford PAD telerobot in the mid 1980's (Cannon and Leifer, 1990a & 1990b; Cannon, 1992 & 1994), is now continuing at The Pennsylvania State University. The purpose of the VR-PAD program is to develop interactive robotic systems where an operator is able to simply point—by manipulating virtual tools in live video scenes—and give object-level directives that specify robotic activities. Using such systems, operators will be able to use robot capabilities previously restricted to highly structured environments in new unstructured tasks such as hazardous waste remediation, space exploration, and flexible manufacturing. Toward the development of such point-and-direct robotics capability, an environment has been created where virtual entities apparently interact with real objects. Unlike purely virtual reality, this mingling of virtual and actual worlds that we call interwoven reality, utilizes live visual information. Virtual tools, such as a graphical representation of a robot gripper, possess latent kinematic and dynamic attributes, but are unencumbered by such properties as they are guided effortlessly to objects in the real scene during robot programming of key points. After an operator uses the virtual tools to specify position and orientation of end-effector key-points, the computer automatically displays a trajectory that accounts for real robot links and dynamic properties. Then the operator simply approves or modifies the trajectory before execution. In our interwoven reality, depth perspective is incorporated to scale each virtual tool so it appears the proper size in all views as the operator moves it around in the live video scene. Correlation points between the live and virtual worlds are established at objects of interest using camera triangulation. Virtual tools appear to disappear behind real objects as if they too are real. A functional VR-PAD system was implemented in the Industrial Engineering Department's Computer Integrated Manufacturing (CIM) Laboratory at the Pennsylvania State University, and is reported here. Various aspects of the system are now being studied.

* Work in these two areas is being funded by the Sandia National Laboratories and The National Science Foundation, respectively.
2. BACKGROUND

2.1 Earlier Point-And-Direct Research

The original Stanford PAD research (Cannon, 1992, ibid.) demonstrated that an operator could direct a robot to perform tasks in a natural and interactive way by pointing to objects and destinations while giving verbal directives such as "put that there" (Figure 1). In this system, both cameras, one above the other, pan together while each tilts separately to specify coordinates of an object by triangulation. The operator remotely aims camera lines of sight to target a location, such as an object, in an unstructured environment. This point-and-direct approach was found to be an order of magnitude faster than using a teach pendant for programming a robot to relocate three objects from various heights above one table to a destination on a second table. Operators also reported that pointing only to destinations entailed less overall fatigue than would be expected for full master-slave telemanipulation of entire trajectories. The concept takes advantage of a robot's basic trajectory planning and auto grasping capability.

![Figure 1. For the Stanford PAD telerobot, a remote operator views two video monitors and aligns objects in each view by rotating the camera on the robot remotely. After the operator points at each location (dashed lines are camera lines of sight), he or she gives directives such as "put that there." The robot then builds a task sequence and trajectory plan for its arm and mobile base to execute. Before sending the robot to perform a task, or anytime during the task, the operator may review the proposed action for changes desired.](image)

In demonstrations, the Stanford PAD Telerobot put items in a tackle box, trash in a wastebasket and blocks on a pallet. These tasks were intended to loosely represent applications in space telerobotics, hazardous waste disposal, and manufacturing material handling. Speed and accuracy was studied and a control theory rationale for deriving the parameters of Fitts' law for new human-machine systems was established (Cannon and Leifer, 1990b, ibid.; Cannon, 1994, ibid.).

2.2 Other Research on Related Systems and Component Technologies

Several lines of research provide additional background to aspects of this work. At one level, VR-PAD robotics is an extension of the basic computer mouse that is used to manipulate icons on a display monitor. Apple Computer Company commercialized this click-and-drag approach. At the Massachusetts Institute of Technology Media Laboratory, Bolt and his colleagues (1984) extended the concept by introducing a wall-sized screen on which icons of naval vessels were manipulated by pointing while saying: "put that there." These approaches were two-dimensional and they were not developed for control of robots.

Other concepts that employ interactive human involvement with certain features that are similar to point-and-direct robotics, though with more continuous path emphasis, more modeling structuredness or on a scale of navigation rather than grasping, include Noyes and Sheridan (1984), Wilcox and Gennery (1987), Schneider and Cannon (1989), Burtnyak and Basran (1991), and Fong, Hines, and Sims; (1994). They have respectively investigated stick-figure robot precursor manipulation, computer aided remote driving, two-armed robotic space docking, graphic pre-positioning for machine vision and augmented mobile robot navigation. The first of these approaches, at the Massachusetts Institute of Technology, addresses the time delay problem in telemanipulation with graphic pre-manipulations. The second, at the Jet Propulsion Laboratory (JPL) in Pasadena, California, accomplishes gross navigation around large obstacles using on-screen pointing. The third, at Stanford University, accomplishes two-handed robotic capture of moving objects by dragging graphic representations of the objects to
destinations in a 2D graphic model. The fourth, by the National Research Council of Canada, allows the human operator to graphically designate locations of interest so that a machine vision system can build a model for use in robotic task planning. The last, in the Information Sciences Division of NASA-AMES which acquired the Stanford PAD Telerobot, is developing systems in which an operator graphically designates a mobile robot course with obstacle avoidance en route. A gesture language is also under development as well as other pictorial communication methods in virtual and real environments (Ellis, 1991). Others including Rocheleau and Crane (1991) as well as Christensen et al. (1991) developed virtual reality models for controlling robots in hazardous waste applications with video toggling as a check. Kim and Bejczy (1991) have implemented video overlay with virtual teleremote for predictive displays. In general, however, virtual reality research in robotics has not addressed primarily at training of operators and virtual telemanipulation. Our point and direct approach avoids telemanipulation by specifying key points, not paths, but similarly implements virtual display and video overlay capabilities.

In regard to instrumented gloves, VPL developed a Data Glove, to manipulate on-screen virtual images (Wise et al., 1987). Their instrumented glove interface let the human operator manipulate virtual images with hand gestures. The Data Glove system uses optical fibers that measure the degree of bending of the user hand joints and was investigated by Tello (1988). A Polhemus magnetic sensor keeps track of the absolute position of the hand with respect to a given source. The Cyberglove, a similar instrumented glove that also uses a Polhemus sensor was developed by Virtual Technologies Inc. of Stanford, California. The Cyberglove uses strain gauge sensors (Virtech, 1991). Both instrumented gloves track finger position and we use the Cyberglove for our gesture interface which allows us to indicate a virtual tool grab and to give gesture directives to our robots (Wang and Cannon, 1993).

Studies of related input devices have considered various factors such as speed, accuracy, mental preparation time, effects of changing control gains, discrete jump control, accessing time, navigation complexity, display size, and a host of additional issues. A few investigators across this range include Card et al. (1978), Buck (1980), Ewing et al. (1986), Fisher et al. (1990) and Traenkle et al. (1991).

Since individual issues of controlling such devices are often hard to isolate from specific technologies, a common practice is simply to compare two or more systems at the current level of commercial capability. This simple comparison is the approach we have taken in our initial pilot test at this early stage. There is, however, a more fundamental foundation for understanding speed and accuracy tradeoffs for many human-machine systems. Fitts' law relates movement time between targets, t_m, to an index of difficulty, I_D (otherwise these terms are given as MT and ID, but subscripts are used here to avoid the misconception of a two variable multiplication in each case). The index of difficulty is a function of the distance, D, that must be traversed between targets and the target width, W. Moving a finger or cursor to a large close target is faster than moving the same to a small distant target. The relationship, including intercept and slope parameters, a and b respectively, is often given as:

\[ t_m = a + b \ln I_D \]

where, \( I_D = \log_2 \left( \frac{2D}{W} \right) \)  

Fitts (1954) developed a rationale behind this relationship from an information theory point of view. The relationship was recently rephrased in terms of the natural logarithm, \( \ln \), in efforts to use control theory to derive the intercept and slope parameters for systems with complex human and machine dynamics (Cannon and Leifer, 1990b; ibid.; Cannon, 1994, ibid.).

\[ t_m = \alpha + \beta \ln \left( \frac{D}{W/2} \right) \]  

where, \( \alpha \) and \( \beta \) are intercept and slope parameters that are possible to calculate before human-machine system construction in terms of step response threshold crossings and dominant system roots from control theory block diagram analysis of known subsystem dynamics. Because dynamic systems generally follow a solution in exponential form for a step response, it was shown that movement time, \( t_m \), can be given as a function of the natural logarithm of the ratio of travel distance to half the target width. Ultimately, it is likely that the parameters of a natural logarithmic relationship for compound movements involving both position and orientation (such as gripper angle during a grasp as well as distance, D) may be predicted using control theory techniques in a similar manner. Future work with our system, now implemented and reported here, will look at this more theoretical issue in a telerobotics context.

3. PURPOSE AND CONFIGURATION OF VIRTUAL TOOLS

The earlier Stanford PAD telerobot was a limited point-and-direct manipulator. While it could be directed to pick up objects and release them at desired locations in unstructured environments, special orientations of the robot gripper could not easily be specified. Because of this, the PAD telerobot was generally configured to approach objects along a line of sight proceeding from where the robot was at the time of pointing with the gripper simply outstretched for a standard grasp. This worked for simple tasks. But sometimes the shape of an object makes it hard to grasp, obstacles exist near the object, or the objects final placement orientation is important. Then, specifying gripper orientation is often needed to designate all aspects of a grasp. Now, incorporating virtual tools, an orientation vector is also specified in a very natural way along with
basic position. By simply orienting our virtual tools using the instrumented glove, one quickly specifies desired positions and angles for a robot end-effector all at the same time (Figures 2 and 3).

Figure 2. Two camera views of a live scene are interwoven with a virtual gripper superimposed on each view in the proper perspective to appear as it would if a real gripper were placed in the workspace. As the instrumented glove moves the virtual gripper in both camera views moves simultaneously. One uses a variety of virtual tools to direct robotic activities in a hazardous waste application. The remote operator effectively shows the robot where and in what orientation to pick up a particular chunk of solid material or cut pipe for example. The robot plans its own path and grasps or cuts autonomously.

Figure 3. An operator grabs and moves a virtual robot gripper to specify gripper orientation as well as position at points of interest for point-and-direct robotic programming. The virtual gripper is interwoven with the video image (only one of the two views is shown). This virtual tool flies in and out of a vertical cut-plane that is defined by a correlation point of camera triangulation corresponding to the depth of a targeted object of interest. The use of wire frame rendering as the virtual tool passes through the cut-plane makes the end-effector seem to disappear behind objects in live video.

The virtual tool shown is an interactive graphical representation of a robot gripper. Additional virtual tools include graphical representations of a sensor suite, cutter, excavator, suctioning tool, machining probe and other robot end-effectors. Distinct from virtual telemunipulation (involving multiple links and joints), a VR-PAD robot calculates its own link trajectories. The virtual tool need only specify end-points of interest. Also, interweaving virtual tools with the real scene reduces the need for world modeling and provides a reality check. Virtual reality, alone without video overlay, can be
convincing but dangerously wrong about obsolete obstacle positions. The live image that is always present in our interwoven reality environment clearly shows, however, when modeled object locations have become obsolete and provides a ready reference for interactively moving virtual objects to where they belong or for otherwise updating the virtual scene. In our interwoven reality, a virtual tool appears as if present in the live scene in several key respects. Besides graphic scaling to insure proper size relative to objects of interest in the scene, virtual tools also partially disappear behind real objects as they begin to pass behind the plane of the object (Figure 3). This effect is accomplished by creating a graphic penetration of a cut-plane. The cut-plane is vertical and normal to a radius from the camera tower at a distance determined by the intersection of the two cameras' lines of sight—the triangulation depth. Live video is superimposed onto this plane at the depth in the virtual workspace corresponding to the triangulation depth of the real object in the real workspace. In the interwoven reality workspace, an algorithm assigns 3-D position to an image object of interest and it is here that we merge the virtual tool graphics with video.

The interwoven reality is created on a Silicon Graphics IRIS Indigo R4000 workstation equipped with an input video board to accept live video views from field based cameras. The virtual tool is slightly transparent so the graphics never completely occlude the real scene. This is accomplished by underlaying the video signal beneath dithered graphics. By using dark colors the dithering effect scatters black pixels liberally throughout the interior of the drawn polygons. The underlay video is configured to overwrite all black pixels, which results in the transparent effect. An algorithm draws the portions of the virtual gripper that are behind a cut-plane as a phantom image so that an object at the depth of the cut-plane appears in the foreground while such obscured portions of the gripper are barely observable behind the object. (The phantom outline remains for reference purposes as an enhancement of reality in which an occluded surface would otherwise be completely hidden from view.)

4. Pilot Test

An initial pilot test was performed with the new system to determine whether investment in an instrumented glove (specifically the attached tracker) that allows simultaneous control of 3D position and orientation is merited for control of virtual tools given that a standard computer mouse can specify the same movement information one degree of freedom at a time using menu options. The pilot test was a simple comparison of two methods of controlling our virtual tools within the newly created interwoven reality.

For this test, a target board with nine random targets was designed to measure targeting performance in a task involving orientations as well as positions (Figure 4). Small light bulbs on the target board were alternately illuminated using a programmable logic controller. Associated with each target light was an adjacent targeting triangle to define orientation. The light bulbs were normally turned off but were sequentially illuminated for the operator to acquire by moving his or her virtual tool until the lighted bulb was engulfed and the two triangles coincident within a 10% alignment threshold. The board was placed on a platform with a CCD camera mounted overhead. The live video scene was projected into a full-screen window on the Silicon Graphics Workstation. The virtual gripper was interwoven with this physical world as described earlier. The two interface device variations were then tested against each other and against a bare hand benchmark where two translations and one rotation were required for acquiring each target.

For the standard mouse approach, an operator controlled each degree of freedom sequentially by first selecting a menu option that determined which degree of freedom the mouse would control. This menu method is one of several possible mouse implementations, and was selected because it could ultimately be expanded to achieve three translations and three rotations for a full six degrees of freedom.

Our instrumented glove configuration converted glove position directly into virtual gripper position and orientation. A subject performed translations and rotations simultaneously by simply moving and reorienting his or her gloved hand in a single motion. A grasp gesture (a closing of the fingers) engaged the glove with the virtual tool so the two would move together in translation and rotation.

The barehanded benchmark, necessary for calibrating the random task to an independent standard, simply involved aligning a Plexiglas gripper surrogate with each target in the same sequence as for the previous two tested conditions.

This test was not a universal comparison of two input devices. Indeed, the mouse cannot be compared as a direct alternative to the instrumented glove in this case because it cannot directly control more than two degrees of freedom simultaneously. Similarly, while the glove alone could be used to test differences between sequential and parallel control of multiple degrees of freedom, it was the relative merit of the glove to a more standard alternative that was of design importance. As a test of whether to invest in an instrumented glove, given that a standard mouse could suffice, the test gives one instance for comparison at the current state of each technology.
Figure 4. Randomly located targets, numbered 1 through 8 consist of a light bulb plus a target triangle. The virtual gripper angles are shown measured counterclockwise from a horizontal line emanating from the target light. The test subject moved the virtual gripper shown to each target light in turn, aligning the orienting triangle of the virtual tool with the target triangle.

Testing of six novice subjects for the three methods (six trial combinations that controlled for cross-test experience) produced subject averaged learning curves with recognizable learning curve features. The three methods tested were catalogued as the menu, glove, and hand methods respectively (Figure 5). Corresponding equations for these subject averages for each of the three methods over the testing range were:

\[
T_{\text{menu}} = 161 - 30.01n \quad (R^2 = 0.99)
\]

\[
T_{\text{glove}} = 106 - 26.8n \quad (R^2 = 0.95)
\]

\[
T_{\text{hand}} = 38.6 - 9.89n \quad (R^2 = 0.99)
\]

where, 
- \(T\) = total eight-target task completion time (in seconds)
- \(n\) = number of trials
- \(R^2\) = level of curve fit (correlation coefficient squared)

After testing the six novice subjects, two additional experienced subjects were tested (Figure 6). These two were tested separately and longitudinally on the system with which they were most familiar over an extended period of time. Performance for these experts was compared for their final 9 of 30 trials. The menu expert averaged 98.4 seconds and the glove expert 37.1 seconds, with standard deviations of 10 seconds and 4.8 seconds, respectively, for the randomly oriented multi-target task. Bare-hand benchmark performance for an expert averaged 20 seconds with a standard deviation of less than one second.
Figure 5. The learning curves for both the menu method and the instrumented glove method for controlling a virtual tool have steep slopes indicating rapid improvement with practice. The lower curve is the barehanded benchmark case of directly moving a physical pointing puck on the test board by hand with no virtual interface.

Figure 6. Experienced operators performed thirty trials for the randomly oriented multi-target task.
The novice subject learning curves are consistent with a premise that the glove is faster than the mouse menu method but not as fast as bare hand manipulation. The extended longitudinal tests for the experienced operators are consistent with a premise that the instrumented glove remains a faster method than the menu method with practice. The data is also consistent with an assumption that learning is still continuing after 30 trials. This was a pilot test and achieving full statistical significance would require additional testing.

A human expenditure ratio can be constructed that is the ratio of the amount of time required to program a robot to do an unstructured task relative to the benchmark time it would take to physically demonstrate the task with one's bare hands (equation 6).

\[
E_{\text{expenditure ratio}} = \frac{T_{\text{robot programming time}}}{T_{\text{baredhanded demonstration time}}} \tag{6}
\]

We calculate this ratio such that an improvement over benchmark produces an expenditure ratio that is less than unity. The time needed to direct a telerobot to complete tasks using virtual reality based point and direct techniques may ultimately take less time than otherwise required for a human operator to physically traverse the area to demonstrate tasks. This is because moving the cameras to specify distant locations in a large workspace (or adjacent facility for a mobile robot) is likely to be faster than physically moving to such locations. We colloquially call this “camera surfing.” In such cases, the expenditure ratio may be less than unity. Other approaches, such as telemanipulation, may have expenditure ratios that are higher than unity for nearly all unstructured tasks since the best possible performance in such cases is achieved (at great expense) when the robot is merely as fast as the human who guides it if unencumbered. Our glove pilot test, reported here, featured no spatial (camera surfing) advantage so all expenditure ratios for the pilot test configurations remain greater than unity as expected.

5. Conclusions

A system that interweaves virtual and physical reality was created in which a virtual tool was successfully interjected into live video and used to specify locations and end-effector orientations for robotic gripping tasks. A cut-plane was demonstrated as a means of relating virtual workspace depth to objects in video on a workstation monitor. A pilot test supported the use of an instrumented glove as an alternative to a mouse based menu method for controlling a virtual tool in an interwoven reality environment. An expenditure ratio was proposed to compare the time needed for programming a robot (e.g. by pointing while giving directives) to a benchmark time needed to demonstrate the same task by hand. As our system expands in capability, and as more virtual robotic tools are developed to strategically designate desired robot end-points in new applications, a means is in place for evaluating alternative systems.

6. Future Work

Virtual Reality based Point-And-Direct telerobotics creates an interwoven reality in which a human operator points to locations using a virtual tool while giving directives by gesturing. In this domain a human operator is able to interact with robots and machines in a natural supervisory manner much like giving instructions to another human.

The VR-PAD system is being extended to include control of multiple robots and multiple tasks. The virtual gripper thus far described is a graphical representation of our custom built robotic gripper for the Puma 560 robot. An Adept robot will soon also be controlled from our Silicon Graphics workstation with video overlay in the same manner. Other virtual tools under development include: cutting, excavating and instrumentation placement tools for hazardous waste remediation; a virtual scribing tool for designating die polishing tasks (Kesavadas and Cannon, 1993); tools for virtual designation of inspection tasks (Wang and Cannon, 1994) and a virtual welding tool with which an operator can point to locations and direct a robot to “weld from there to there.” In the planned implementations, these virtual tools will be stored as icons in a virtual toolbox. The operator then may select and use these tools for various portions of a task sequence. The selection of a virtual tool icon directs the robot to enact a corresponding tool change in the program when it is executed. With respect to formal treatment of the human-interface, Fitts' law equations will be developed from control theory for compound movements including both position and orientation. Future tests will involve three-dimensional out-of-plane positions and orientations. Demonstrations will also include directing robots to perform additional tasks in hazardous waste remediation, space telerobotics and flexible manufacturing. In such tests, expenditure ratios for developing robotic tasks using point-and-direct virtual tools will be compared to teach-pendant programming and various types of telemanipulation. In the end we hope to demonstrate the best of both physical and virtual worlds in enhancing our interactions with robots and machines.
7. References


