Proc. American Nuclear Society (ANS) 7th Topical Meeting on Robotics and Remote Systems
Augusta, Georgia, USA, April 27-May 1, 1997. pp. 966-973.

AN AUGMENTED REALITY BASED TELEOPERATION INTERFACE FOR UNSTRUCTURED ENVIRONMENTS

Paul Milgram
Mechanical & Industrial Engineering,
University of Toronto
Toronto, Canada M5S 3G9
(416) 978-36662
milgram@mie.utoronto.ca

Shi Yin
Mechanical & Industrial Engineering,
University of Toronto
Toronto, Canada M5S 3G9
(416) 978-36662
milgram@mie.utoronto.ca

Julius J. Grodski
Defence & Civil Institute of
Environmental Medicine (DCIEM)
Downsview, Ontario, Canada M3M3B9
(416) 635-2085
jul@dciem.dnd.ca

ABSTRACT

The general problem of managing a remotely situated vehicle or manipulator system is discussed, from the point of view of what level of autonomy is feasible. Assuming continued presence of a human operator (HO) in the loop, the advantages of Director / Agent (D/A) control are discussed, as a means of alleviating the tedium of conventional continuous manual teleoperation. In order to realise a viable D/A system, the HO must be able to communicate accurate quantitative 3D task related data to the robot controller. In an unstructured environment, such information is typically neither available a priori nor readily obtainable during task execution. An Augmented Reality display system is introduced as a means to obtain such quantitative measurement data on-line, via *partial modelling* of the remote site using a *Virtual Tape Measure*. Some of the design considerations for this system are discussed and sample measurement accuracy and precision data are presented.

I. INTRODUCTION: DIRECTOR/AGENT CONTROL FRAMEWORK

In a general sense, the control of any remotely located vehicle or manipulator typically comprises a wide assortment of activities, to be carried out at various times and under varying circumstances. Included among the basic functions underlying those activities are the following:

- deciding which objects at the remote worksite are to be operated upon;
- deciding which objects at the remote worksite are not to be operated upon, i.e. avoided;
- ascertaining the locations of such objects, as appropriate;
- deciding how and where they must be manipulated or relocated, as appropriate;
- planning the path of the remote vehicle and/or manipulator;
- communicating command information from the control centre to system actuators.

As the sophistication of techniques for managing tele-robotic systems continues to grow, thereby enabling their deployment in increasingly more remote and more hazardous environments, it is nevertheless clear to those familiar with the realities of even the most advanced control technologies that complex robotic tasks are unlikely to be achievable at the high levels of autonomy once envisaged, and especially not in highly unstructured and dynamically varying environments. It is evident, in other words, that, due to cost constraints, technological limitations, and/or operational uncertainty, some form of mediation by a human operator (HO) will be required for some time to come.

Although complete autonomy may be an elusive goal, this does not imply that the tedium of continual, sustained manual involvement by human operators can not be significantly alleviated. In Fig. 1 we depict the contrast between manual versus autonomous control in a simplified fashion, not as a dichotomy but as a *continuum* spanning a range of levels of human involvement in remote control. Manual teleoperation, at the bottom of the spectrum, encompasses all situations in which the HO is constrained to remain *continuously* within the control loop. In other words, if the HO stops controlling, the loop is opened. Note that the otherwise technologically quite sophisticated concept of *tele-presence* control is depicted somewhat ironically also at the bottom of the Fig. 1 spectrum, as a special case of manual control. From the point of view of human involvement, telepresence is taken here to comprise any system for which the HO is made to feel that the remote effector is part of herself and thus that she is effectively present at the remote site. Typically, this involves some

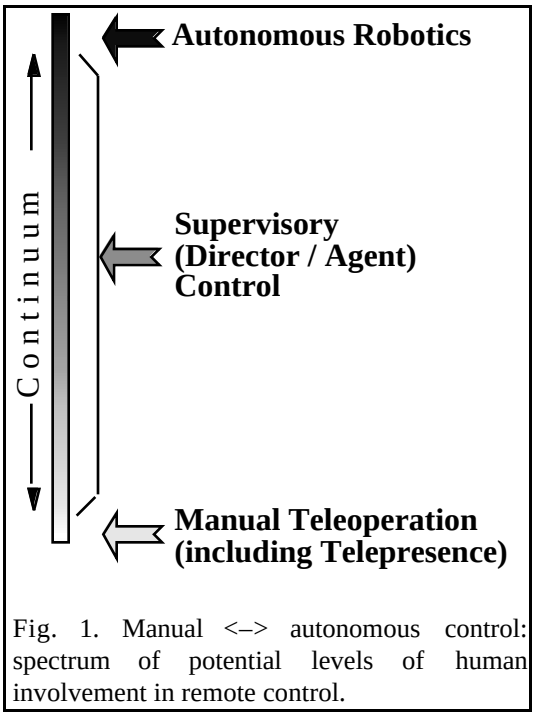

Fig. 1. Manual <–> autonomous control: spectrum of potential levels of human involvement in remote control.

form of master-slave control, where all actions of the master arm initiated by the HO are mimicked by the slave manipulator. Clearly, therefore, in terms of the autonomy spectrum in Fig. 1, this is essentially equivalent to manual control, since whenever the HO ceases to direct remote movements, the loop is effectively opened. Conversely, remote autonomy and immersive master-slave telepresence are, for the same reason, incompatible concepts.

From the point of view of control, the manual teleoperation concept is depicted in the bottom part of Fig. 2. Because the loop is always closed, the default strategy is for all commands exiting the control interface to be sent directly to the remote manipulator or vehicle. In many cases, this approach is liable to result in a collection of clumsy "move-and-wait" or exploratory "poke and probe" manoeuvres. This is especially critical for situations which are characterised by instabilities due to time delays, and/or by degraded images due to limited bandwidth communication channels, and/or by the presence of volatile or hazardous substances with which inadvertent contact is to be avoided at all costs. If, on the other hand, control can be effected in an *off-line* fashion, by providing the HO with a means to plan, program and rehearse eventual control commands before they are actually sent to the remote system for execution, many of the problems listed above may potentially be averted.

Such is the principle underlying the region spanning the region between the two extrema of the continuum in Fig. 1. Here we have employed the label "Supervisory Control", whereby the HO assumes the role, in a general sense, of managing the operation of a collection of variously autonomous system functions[1]. "Director / Agent" (D/A) control, indicated in parentheses, is considered here as a basic form of supervisory control, where the HO acts as a *director* and the limited intelligence robot acts as her *agent*[2,3]. One manifestation of such a scheme might entail the Director indicating the location of an object to be grasped and picked up, whereupon the locate-grasp-and-pick operation would be carried out independently by the Agent at a later time. With this approach it is therefore not necessary for the HO to feel *present* at the remote worksite, but may rather feel herself *adjacent* to the robot. In addition to alleviating the need to remain continuously in control (which might in fact become rather tedious for tasks which are either lengthy or which execute at uncomfortably slow rates), D/A control also places less stringent technological requirements on the design of the human/robot (H/R) interface. That is, not only is high fidelity master-slave control hardware not essential according to this concept, but any requirement for a telepresent HO to be immersed in the remotely viewed environment, by means of a head-mounted display (HMD) system for example, is also obviated. Instead, a much simpler *monitor-based* display system, with which the HO effectively *looks in at* the remote world, may suffice quite well.

II. PARTIAL MODELLING OF REMOTE SITE

The key capability needed for D/A control to be feasible is a means by which the HO can accurately and reliably communicate the necessary instructions to the robot for subsequent execution. The general means of realising this is shown in the top part of Fig. 2, where the HO ideally has available a model of both the remote worksite and the remote robotic manipulator. In other words, rather than manipulating the real system, as in on-line control, the HO is now able to manipulate a quantitative 3D *model* of the remote system. This then provides the opportunity for the HO to try out candidate control manoeuvres in non-real time and release the final commands for execution only when she is convinced that the planned manoeuvres can be carried out safely and satisfactorily.

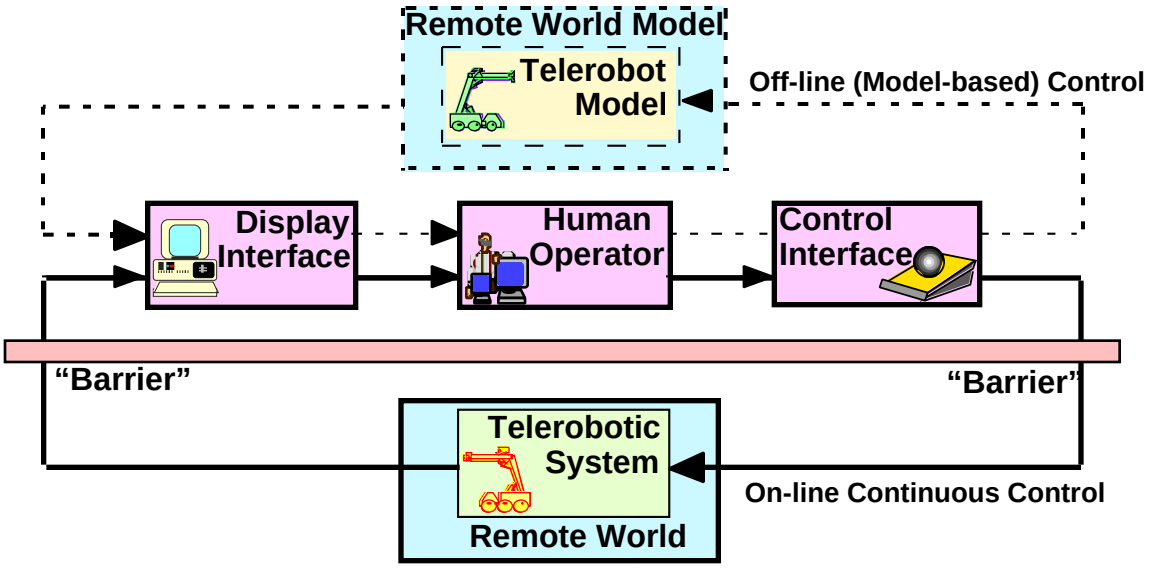

Fig. 2. Generic telerobotic control system. Bottom looop: closed loop, continuous, direct control of real system. Top loop: off-line control via simulation model.

The principal challenge in realising such a control scheme is the acquisition of sufficient information about the remote manipulator/vehicle and the remote task space to enable construction of adequate models. Clearly some models will be better than others, however, and thus once again it is convenient to consider a continuum of such models, as illustrated in Fig. 3. At one extreme we have the case of the remote world being fully modelled. This refers to situations in which the robotic environment and the objects being manipulated within it are largely invariant and are well known in advance, and the operational procedures are prespecified. Such a case would describe many industrial robotic applications, for example, in which the robots, the worksite and the parts being manipulated have presumably been modelled using CAD and fabricated to a known degree of precision. The opposite end of the spectrum in Fig. 3 refers to remote environments which are completely unknown, that is, usually unstructured environments for which no prior knowledge is available, thereby making that world very difficult to model.

The continuum of intermediate cases illustrated in Fig. 3 covers a variety of situations in which some knowledge of the remote world has been obtained. For example, prior knowledge might be had about the sizes, shapes, etc. of objects to be manipulated, while nothing is known about *where* these objects are located. Alternatively, particular object locations might be known or measured, while nothing is known about *what* the associated objects actually are, or what relevance they might have to an operation.

Referring to Fig. 1 and the concept of Director / Agent control, we recall that the key requirement for allowing the HO to operate as more than just a continuous manual controller is the ability to acquire adequate quantitative information about the remote world to initiate subsequent semi-automated machine execution. For environ-

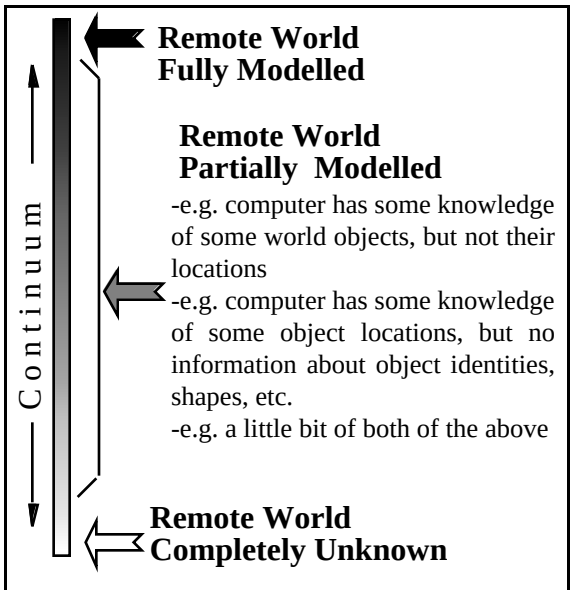

Fig. 3: Spectrum of levels of knowledge about remote worksite and manipulation system.

ments which are not completely modelled, the challenge therefore is to provide the HO with this information. One way to solve this problem is to use whatever means are feasible to construct as complete a model as is possible, by making use of laser scanning or machine vision systems for example. Our own somewhat contrasting approach is to suffice with a much sparser database of quantitative measurements, by interactively building up a sufficient *partial model* of the worksite.

The concept of interactive partial modelling is proposed, in other words, as a means to exploit the presence of an intelligent human operator (HO) to convey limited amounts of data about the video scene to the computer, and thereby to develop and refine a quantitative, but not necessarily complete, model of portions of the remote world. On the basis of this partial model, the spatial information necessary for defining targets, waypoints and trajectories, as well as boundaries and obstacles, can be communicated for subsequent semi-automated control of the telerobot.

III ARGOS: VIRTUAL TAPE MEASURE

We have successfully implemented the essential cornerstone of an interactive partial modelling system in the form of a *virtual tape measure*, which is capable of making accurate absolute 3D measurements of object locations in a remotely viewed video scene. This is accomplished by applying the *ARGOS* (Augmented Reality through Graphic Overlays on Stereo-video) toolkit, developed in our lab [4,5,6]. ARGOS is a "Mixed Reality" display interface[7] employing calibrated stereoscopic 3D graphical overlays on a remote stereoscopic video view of the real (unmodelled) world. The major components of ARGOS technology are stereoscopic video and calibrated stereographics. These are discussed below, with emphasis on their applicability for teleoperation.

The tools provided by ARGOS can be classified into "probing tools" and "enhancement tools". All have been designed both for improving the HO's comprehension of the remote environment and for *interactive modelling* of the remote world.

• The most basic probing tool is the *virtual pointer*. This is a stereo-graphic cursor which can be positioned anywhere in the stereo-video scene. When properly calibrated, the virtual pointer gives a direct readout of its corresponding {x,y,z} location in absolute real world units, and thus quantifies the 3D location of any object adjacent to which it is placed.

• The *virtual tape measure* is an extension of the virtual pointer, used for measuring distances between points in the remote stereo video scene. It is generated by clicking a start point with the virtual pointer and dragging a virtual line of calibrated length through the video image to a selected end point.

• Related to both of the above are *virtual landmarks*, which are graphical objects of known length, or known separation, superimposed on the video scene to enhance the HO's ability to judge absolute distances, and thus the absolute scale of the remote world. Research has shown that such simple graphic aids can be very effective for this purpose[8].

• *Virtual planes* are generated by specifying three or more coplanar points with the virtual pointer. One important application of such planes is for restricting movement of simulated objects. For example, a graphical model of a (virtual) robot interactively placed on a real surface would not ordinarily have any knowledge about the planar constraints of that surface. A straightforward way of conveying such information to the computer would be through interactive modelling using virtual planes.

• *Virtual objects*, which are either interactively generated or premodelled according to particular geometric specifications, can be superimposed on stereo video at designated locations and at specified orientations to appear as if they are really present within the remote scene.

• *Virtual encapsulators* are wireframe shapes created on the remote stereo video scene to encapsulate real objects. This can be done approximately, as a tool for indicating an envelope of size, position and orientation of a real object in space, or more exactly, for highlighting the edges of an object. Virtual encapsulators require the same modelling, location and orientation data as do virtual objects.

• *Virtual trajectories* are graphical indications of prescribed robot motions, added to the image of the real robot at a particular initial configuration, to specify the desired trajectory for the robot to follow. These can be used, for example, for path planning purposes, by placing trajectories into the video space and verifying plans for their accuracy in relation to the actual (unmodelled) worksite.

• A subclass of the virtual trajectory is the *virtual tether*, which was developed as a perceptual aid for manual telemanipulation tasks. A virtual line, or tether, is drawn between the end effector and its intended target. As the manipulator moves, it remains "tied" to the target through the tether. Research has shown that use of such a tether was able to improve accuracy in an experimental peg-in-hole task, relative to the case of stereo video alone with no tether[9].

• Finally, a full stereographic 3D model of a remotely controlled robot, or a *virtual manipulator*, has been

developed. By superimposing such a model of the robot at the remote site onto the real robot, the graphical model can be manipulated within the real (complex, unstructured, unmodelled) 3D work space. This concept is discussed further below.

The common feature in all of the above is the virtual tape measure (VTM) as the basic measuring instrument. To use this, a coordinate point in the stereo video scene is defined by moving the overlaid 3D stereographic (virtual) pointer in 3D stereo video space to match the perceived 3D location of any selected object feature. In order to transform this mapping of the two images into a real world measurement, it is essential that a unique bidirectional one-to-one mapping of coordinate spaces be established between the virtual graphic world and the remote world viewed through stereo video. In the ARGOS system calibration and registration of graphics and the real world is accomplished[10] by means of a calibration object of known dimensions situated within the video image. (For teleoperation applications the robot itself can serve as the calibration object, since it is in any case present at the remote scene and its geometric parameters are assumed to be well known.)

Two modes of operation of the ARGOS virtual tape measure (VTM) have been implemented to date. In the basic *unaided* mode, the accuracy and precision of measurement rely solely on the visual perceptual capabilities of the operator, who must align the virtual pointer as described above based on where he perceives the pointer to be in relation to the object feature whose 3D location is being measured. In the *aided mode* we have added a second stage, whereby the user is able to avail himself of a second toolbox of algorithms which attempt to lock the pointer automatically onto a particular class of image feature located within the predetermined immediate 3D vicinity of the pointer.

The aided alignment feature, which continues to be developed to encompass a library of specialised image feature extraction algorithms (a notion which is feasible for real remote manipulation operations due to the fact that a human remains in the loop to determine which features are relevant for any particular video image), currently comprises only corner (junction) characterisation[11]. It operates by characterising a corner in grey-level images by the number of lines and orientations which constitute the corner. A corner can therefore be classified as an L-junction, a T-junction, or a Y-junction. In principle, the corner location (intersection point of the lines) can also be detected with subpixel accuracy. The approach is based on the statistical analysis of gradient-direction in an intensity image, and takes the signal-to-noise ratio (SNR) into account.

IV. ACCURACY OF VIRTUAL TAPE MEASURE

The obvious critical question to ask about the VTM is how precise and how accurate are its measurements. In terms of precision, there are several factors which will determine this performance measure:

- the separation and alignment of the stereo video cameras, which together determine the magnitude of the resultant horizontal image disparity, which then translates into amount of resolvable depth;
- the focal lengths of the camera lenses, which similarly determine the size of the image disparity for corresponding differences in the real world being imaged;
- the resolution of the cameras, the video board and the monitor;
- the ability to overcome the discrete nature of the display devices and achieve subpixel accuracy;
- ambient lighting;
- the quality of the 3 DOF control device for manipulating the cursor.

In terms of accuracy, there are four primary factors which can contribute to measurement error:

- errors due to the calibration procedure;
- errors due to optical lens distortion, and whether or not such distortions have been compensated for;
- errors introduced by the user (presumably inadvertently) in aligning the cursor with the target feature;
- errors due to the feature detection subsystem.

Initial studies with the VTM in the unaided mode have shown that subjects were able accurately to align the position of a virtual graphic pointer relative to the positions of real video targets as well as they could when a real pointer was similarly manipulated in the real video scene[12].

As a result of the many factors listed above which contribute to the precision and accuracy of an operating VTM system, it is clear that performance is highly dependent on specific implementation parameters, as well as operator skill. For example, if it were to be found that measurement accuracy for a particular setup was unsatisfactory, it would be quite rational to consider modifying it to incorporate a larger camera separation, or larger focal length lenses.

Nevertheless, a small experiment has been carried out to provide the reader with a sense of what sort of performance figures might be expected, from one particular set up at least. The experiment was performed

using a pair of Hitachi VK-C150 CCD cameras, which were mounted with a separation of 12.5 cm and were set to converge at a distance of 74 cm. It is important to note that this particular setup was used out of convenience and therefore that the data generated reflect the scale of this setup. One can therefore reasonably expect similar *relative* accuracy for more distant objects, by simply scaling up the camera system accordingly. Of course, even better performance data could be achieved, by invoking some of the factors mentioned above, such as larger camera separation.

In the experiment, three targets were set up, to form the sides of an imaginary triangle in 3D space. The distances comprising these sides were: A) 17.3 cm; B) 19.0 cm; C) 25.1 cm. Fig. 4 shows the experimental setup which was used in the experiment, with the three target points slightly highlighted. The particular separations were chosen in order to fill up the screen as much as possible. Six experienced volunteer subjects were used, to make 18 measurements each. These comprised two measurements each (aided and unaided) for each of the three target distances. The entire target setup was also moved back and forth, and the measurements were repeated for three different mean distances from the centre of the cameras: Near = 64 cm; Mid = Convergence Distance = 74 cm; and Far = 90 cm. (In other words, the number of runs per subject = 3 separations x 3 target distances x 2 aiding conditions = 18.)

The cumulative results of the experiment are shown in Fig. 5. Rather than computing mean values, we have instead elected to show the raw data, to provide a strong visual indication of the consistency of the measurements. As indicated, the right column shows the results of the measurements without the automatic corner detection tool (i.e. based on human perception alone), while the left column shows the aided results. Each column is divided into three sets of graphs, for the three distances from camera to target setup: {64, 74, 90 cm}. Each set of graphs shows three sets of measurement data, labelled A,B,C, which correspond to the three sides if the imaginary measurement triangle. The *actual* distances are placed according to the scale on the horizontal axis, and the measured data are plotted vertically. The diagonal line in each graph represents perfect performance; that is, any datum lying on that line represents a zero error score. The six data plotted for each condition represent the raw measurements of the six subjects.

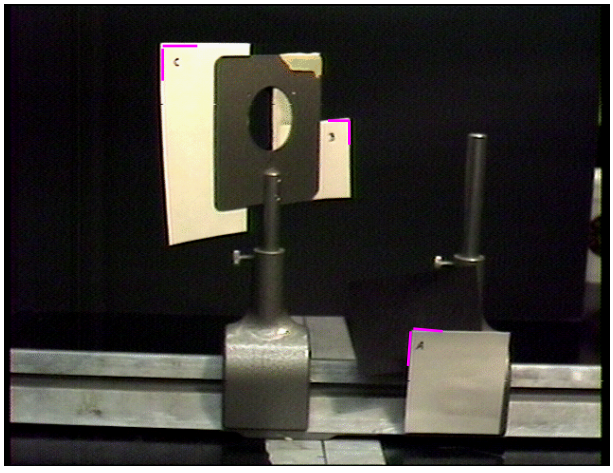

Fig. 4: Setup of accuracy assessment experiment. The slightly highlighted outside corners of the three white rectangles were the targets.

Above each of the raw measurement graphs is a bar graph which shows the mean percent error for that particular set of measurements, averaged over subjects. Each bar graph height represents | measured minus actual distance | / distance x 100. The bars correspond ordinally to the distances A,B,C in the horizontal direction, but are not plotted exactly to scale. Note that the vertical scales for these graphs have a maximum value of only 7.5%.

The general conclusion to be drawn from these graphs is that performance was generally very good, with a few exceptions. First of all, the spread of data in the left column is less than that in the right, indicating that performance with the aided cursor was more consistent, i.e. more *precise*, as was expected from a tool that was designed to make the measurements more reliable. There is also an indication of a slight but consistent positive error; i.e. an overestimation of the measured distances. It is our belief that this is due to the optical distortions in the lenses of the cameras, which have not yet been taken into account in our calibration and measurement procedures. Due to the paucity of data in the experiment, we have not felt it worthwhile to perform detailed statistical analysis on these data. In evaluating the magnitudes of the measurement errors between the left and right columns, there does not appear to be a large difference in accuracy between the two cases, except (surprisingly) for the 74 cm condition in the middle.

In drawing the provisional conclusion that, while *precision* appears to be better with the aided cursor, the *accuracy* appears to be comparable, we also take note of the actual error magnitude plots, which for many will be the most significant results. In general, it appears that we are able to obtain an accuracy of about 3%-5% in our measurements, *with the present system, with this*

particular setup. As discussed above, significant improvements are to be expected if major changes were to be made to the camera alignment parameters and the focal lengths used, and if optical distortion were taken into account.

V. VIRTUAL TELEROBOTIC CONTROL

This concept of interactive partial modelling using superimposed stereoscopic computer graphics forms the basis of the ARTEMIS (Augmented Reality Telemanipulation Interface System) interface described elsewhere.[10,13,14] Because it is often impossible to update the video image from the remote site on a continuous basis, ARTEMIS grabs a single stereo video image transmitted from the remote scene and the operator then uses the local computer to create a partial model of the remote site. Once a sufficient model has been obtained, the desired manoeuvres are conveyed to the control computer and rehearsed. Note that such local off-line control is essentially closed loop, since there is no time delay due to communication with the remote site. Once the manoeuvre has been approved, the operator simply relays the trajectory end point and robot control signals for execution at the remote site. Because the information transmitted at this point requires a very low bandwidth, subsequent execution of the manoeuvre at the remote site can commence effectively instantaneously. ARTEMIS has been tested by controlling our robot at the University of Toronto from several locations around the world, the most distant being from Kyoto, Japan.

ACKNOWLEDGEMENTS

This research has been supported by the Defence and Civil Institute of Environmental Medicine (DCIEM), the Manufacturing Research Corporation of Ontario (MRCO), and the Institute of Robotics and Intelligent Systems (IRIS). The authors also gratefully acknowledge the valuable contributions of Messrs. Kitman Cheung, Ludovic Canas, David Drascic, Douglas Liversidge, Steven Ma and Maurice Masliah.

REFERENCES

1. T.B. Sheridan, *Telerobotics, Automation and Human Supervisory Control,* MIT Press, Cambridge, MA (1992).
2. S. Zhai and P. Milgram, “Human-robot synergism and virtual telerobotic control”, *Proc. Annual Meeting of Human Factors Association of Canada*, (1992).
3. S. Zhai and P. Milgram, “A telerobotic virtual control system”, *Proc. SPIE 1612, Cooperative Intelligent Robotics in Space II*, Boston (1991).
4. P. Milgram, D. Drascic, and J.J. Grodski, “Enhancement of 3-D video displays by means of superimposed stereo-graphics”, *Proc. Human Factors Society 35th Annual Meeting*, San Francisco, (1991).
5. P. Milgram, S. Zhai, D. Drascic and J.J. Grodski, "Applications of augmented reality for human-robot communication", *Proc. IEEE/RSJ Int'l Conf. on Intelligent Robots & Systems (IROS)*, Yokohama (1993).
6. P. Milgram, H. Takemura, A. Utsumi and F. Kishino, “Augmented Reality: A class of displays on the reality-virtuality continuum”, *Proc. SPIE Telemanipulator and Telepresence Technologies*, Vol 2351, Boston (1994).
7. P. Milgram and F. Kishino, “A taxonomy of mixed reality visual displays”, *IEICE Trans. on Information Systems, special issue on Networked Reality* (1994).
8. P. Milgram and M. Krüger, "Adaptation effects in stereo due to on-line changes in camera configuration", *Proc. SPIE Vol 1669-13, Stereoscopic Displays and Applications* (1992).
9. K. Ruffo and P. Milgram, "Effect of stereographic + stereovideo "tether" enhancement for a peg-in-hole task", *Proc. IEEE Annual Conf. on Systems, Man & Cybernetics* (1992).
10. A. Rastogi, "Design of an interface for tele-operation in unstructured environments using augmented reality displays", Unpublished MASc dissertation, University of Toronto (1996).
11. Yin, "An investigation of image capture, compression and feature extraction algorithms for an underwater telerobotics system", Unpublished PhD thesis, University of Trondheim, Norway (1993).
12. D. Drascic and P. Milgram, "Positioning accuracy of a virtual stereographic pointer in a real stereoscopic video world". *Proc. SPIE 1457, Stereoscopic Displays and Applications II* (1991).
13. P. Milgram, "Using augmented reality for telerobotic control", *Advanced Imaging*, (1996).
14. A. Rastogi, P. Milgram, D. Drascic, and J.J. Grodski, "Telerobotic control with stereoscopic augmented reality", *SPIE Vol. 2653 Stereoscopic Displays and Applications VII & Virtual Reality Systems III*, (1996).

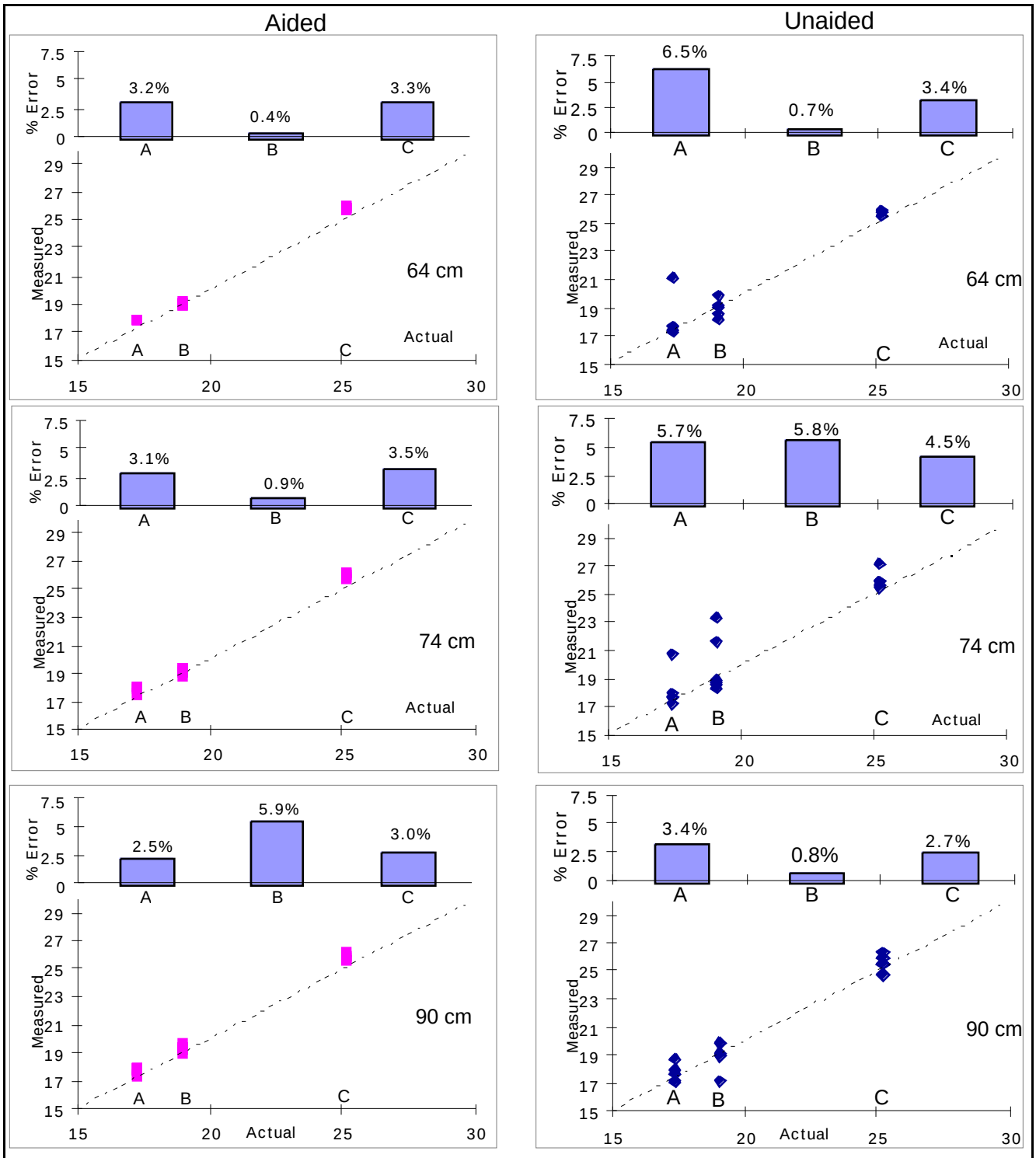

Fig. 5: Measurement data from experimental test of precision and accuracy of Virtual Tape Measure. Left and right columns correspond to Aided and Unaided measurements. Three vertical plots are for different distances from centre of stereo cameras. Each group shows measured vs actual lengths, as well as percent error, for three different target separations.