Telerobotic Control with Stereoscopic Augmented Reality

Anu Rastogi, Paul Milgram, David Drascic*
Ergonomics in Teleoperation and Control Lab., University of Toronto
Email: anu@argos.rose.utoronto.ca, milgram@mie.utoronto.ca, david.drascic@utoronto.ca, ,
4 Taddle Creek Road, Toronto, Ontario M5S 3G9

Julius J. Grodski
Defence and Civil Institute of Environmental Medicine,
North York, Ontario M3M 3B9
Email:jul@dciem.dnd.ca

Keywords: Telerobotics, augmented reality, stereoscopic displays, teleoperation.

1. Abstract

Teleoperation in unstructured environments is conventionally restricted to direct manual control of the robot. Under such circumstances operator performance can be affected by inadequate visual feedback from the remote site, caused by, for example, limitations in the bandwidth of the communication channel. This paper introduces ARTEMIS (**A**ugmented **R**eality **TE**le**M**anipulation **I**nterface **S**ystem), a new display interface for enabling local teleoperation task simulation. An important feature of the interface is that the display can be generated in the absence of a model of the remote operating site. The display consists of a stereographical model of the robot overlaid on real stereovideo images from the remote site. This stereographical robot is used to simulate manipulation with respect to objects visible in the stereovideo image, following which sequences of robot control instructions can be transmitted to the remote site. In the present system, the update rate of video images can be very low, since continuous feedback is no longer needed for direct manual control of the robot. Several features of the system are presented and its advantages discussed, together with an illustrative example of a pick-and-place task.

This paper was published in:
SPIE Volume 2653: Stereoscopic Displays and Virtual Reality Systems III
Editors: Mark T. Bolas, Scott S. Fisher, John O. Merritt
San Jose, California, USA, January - February 1996
pp 115-122

---

2. Background: Conventional Teleoperation Control

In typical cases of manual control of a robot at a remote site, the human operator (HO) communicates with the robot through local control inputs, while a video camera at the remote site sends back information which reflects the result of the HO's control actions. This method of control has a number of drawbacks, however. First of all, the HO is forced to remain continually engaged in the dynamic control of the robot whenever any actions are to be executed. Secondly, because any small delay in the feedback of HO control actions can cause teleoperation performance to deteriorate significantly, it is highly desirable to maintain a relatively large bandwidth in the communication link joining the HO and the remote robot. Unfortunately, this situation is not always feasible, and can have associated with it a significant cost (e.g. expensive satellite links) and/or a large operational impact (e.g. costly and cumbersome tethers on underwater ROV's). Another set of consequences related to the HO's being continuously occupied in the telerobot control loop is the fact that the HO himself might retard the system, due to relatively slow sensory and motor capabilities.

More common, however, is the converse case, where the robot dynamics are very slow, and the HO is compelled to contend with the tedium of having to slow down while remaining tied into the control loop. Finally, there are typically no means included with such manual control system to provide protection against control errors (e.g. bumping into critical objects) committed by the HO while issuing on-line commands.

A further problem associated with most telerobotic scenarios is that operations must typically be performed in relatively *unstructured* environments. Examples of unstructured environments include applications in mining, bomb disposal, underwater inspection, maintenance, etc. Since unstructured environments are by definition difficult to characterise, quantitatively modelling such environments may be unfeasible. In other words, either prior knowledge about the site of operations is insufficient or required properties of the particular objects to be operated upon are not known. An additional, perhaps more serious obstacle, is that unstructured environments are often subject to unpredictable changes in the environment, which therefore inhibits the ability to initiate repetitive programmed procedures and results in a need for higher flexibility in operational procedures.

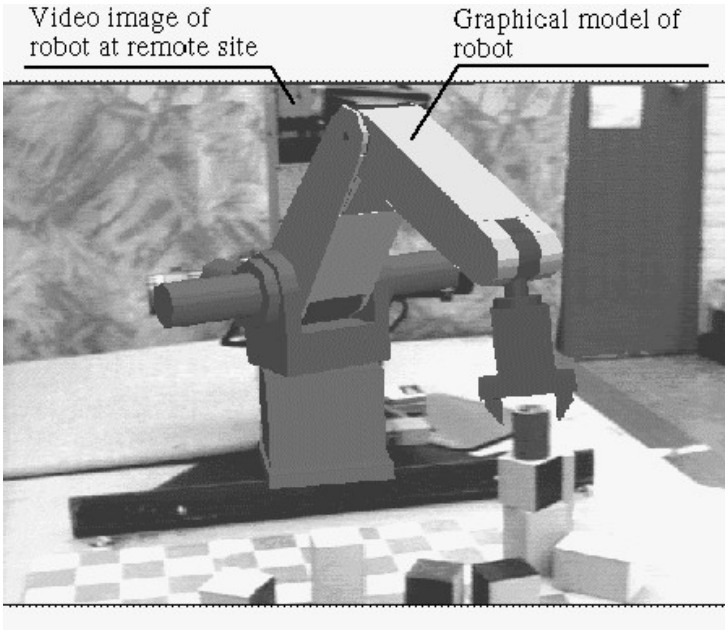

FIGURE 1. *Calibrated and registered graphical robot overlaid on video image from remote site. In the figure, the entire image is video, except for the graphical robot. (Note that all composite stereoscopic graphical overlay and stereovideo images presented in this paper are of necessity presented monoscopically, but should in fact be envisaged according to their actual stereoscopic rendering on the display monitor.)*

It would appear, therefore, that continuous closed loop manual control is the only conceivable option for teleoperation in unstructured environments. Nevertheless, if it were possible to close the operator's control loop off-line, at the *local* site rather than through the entire loop via the remote site, some of the disadvantages of high bandwidth and delays in visual feedback could be reduced or eliminated. Achieving this, however, would require the HO to be able to make use of information that is in fact *not yet* available. In other words, direct operational *feedback* would be insufficient; the HO would have to have *predictive* information available. Because it is impossible to perform any sort of prediction without some knowledge of the system and the circumstances being extrapolated, the key to being able to perform off-line control is having an *adequate model* of the system being controlled. If such a world model were available, local (visual) feedback could be provided by means of a graphical simulation of the task. The operator could then operate the robot within the graphical simulation model and the instructions could then be transmitted to the actual telerobot at the remote site.

In addressing the challenge of supporting teleoperation in unstructured, and thus difficult to model, environments, this paper discusses a new display concept, in which graphical simulation of the telerobotic task can be performed, *even though a complete model of the remote task space is not available.*

3. ARGOS: Augmented Reality Display System

3.1. Basic Concept

In section 4. we will describe detail of ARTEMIS, an **A**ugmented **R**eality **TE**le**M**anipulation **I**nterface System, which has been developed for the purpose of facilitating teleoperation in the absence of an adequate model of the remote world. The key technology underlying ARTEMIS is the use of Augmented Reality, comprising stereoscopic computer graphics combined with stereoscopic video, to display both real and graphically simulated data on a single monitor. Before going into the details of ARTEMIS, we first describe the concept of Augmented Reality as used here. We then discuss a number of aspects of the stereoscopic display system used for the task simulation.

As introduced above, the basic challenge is how, in the absence of modelled data of the remote site, can the desired graphical task simulation be created? Because it is reasonable to assume that a (geometrical and kinematic) model of our own robot at the remote site is in fact available, ideally one would want to use this model to perform simulated manipulations of other modelled (i.e. virtual) objects at the remote site. Since this can not be done, our alternative approach is to use the graphical robot model to "interact" with the unmodelled (i.e. real) objects which the operator is able to see and interpret within the live video image. The resulting *combination of graphics and video* thus serves as a hybrid visual display for off-line task simulation. The display is created by overlaying three dimensional (3D) graphical objects on top of live (unprocessed) video images. Because the net result is an augmentation of the real (live) video data, this method of presentation, by definition, falls within the class of *Augmented Reality* (AR) display7 and forms the basis of the ARGOS™ (**A**ugmented **R**eality through **G**raphic **O**verlays on **S**tereovideo) display system developed in our laboratory4,5,6,9.

3.2. Stereoscopic Display Advantages

A critical property of the display approach introduced above is that both the graphic and video components of the system are presented *stereoscopically.* This key feature plays a role not only for enhancing the HO's perception of the remote site but also for enabling the quantitative measuring capability described below. Although stereoscopic displays can be realised in various ways, they are all based primarily on retinal disparity cues, i.e. differences due to different projections of images onto the left and right retinas. The stereoscopic display used for the AR teleoperation interface presented here exploits these cues to accomplish three unique objectives, which are not otherwise achieved with monoscopic displays: 1) to improve depth perception, 2) to facilitate integration of depth information from both graphics and video, and 3) to enable interactive acquisition of coordinate information from the remote site. These functions are discussed below.

3.3. Improved Depth Perception

One important problem associated with monoscopic (single camera view) pictures such as that illustrated in Fig. 1 is that, even when such composite images are displayed with correct geometrical perspective (see 3.6), the perception of depth can be ambiguous. For example, in Fig. 1 the magnitude of the relative distance in the depth plane between the graphic robot gripper and the real (video) cylinder in the vicinity of the graphic gripper is uncertain. By using stereoscopic displays for both the video and modelled graphic images, this depth ambiguity is reduced.

Stereoscopic video (SV) on its own has been shown repeatedly in earlier studies to improve teleoperation performance. For example, using a simulated teleoperated explosive ordnance disposal task, Drascic3 compared performance measures between monoscopic video (MV) and SV displays. In addition to finding significant benefits from the use of SV displays for the more difficult tasks, he also found that for the simpler tasks SV showed better performance than MV during the early stages of skill acquisition. Draper, Handel and Hood2 compared SV with MV for a Fitts' Law tapping task and also found advantages with SV, especially for higher indices of task difficulty and for inexperienced subjects. In another experiment, with reduced perspective and shadow cues in the display, they again found significant advantages with SV. In that paper, Draper et al concluded that teleoperation performance improves with SV when image clarity is reduced, when the task is not structured, and when skill levels, experience and dexterity are low.

3.4. Combination of Graphic and Video Depth Information

Human perception of depth in real world images is based on a variety of depth cues. (For a detailed overview refer to Boff, Kaufman and Thomas1) When viewing monoscopic video images, for example, objects in front of other objects occlude those objects situated farther away (interposition cue); some objects cast shadows on other objects, depending on the lighting conditions at the remote site (light and shadow cue); parallel lines tend to converge at greater distances (linear perspective); and the sizes of familiar objects appear smaller whenever they are farther away (size constancy cue). In recognition of the importance of these cues in real-life images, similar effects are often simulated when generating 3D computer graphic images.

However, whenever graphical objects are overlaid on video images for AR displays, some of the 3D depth cues incorporated into the graphical objects may in fact not be consistent with the depth cues of the corresponding video image. For example, the graphic objects may not obey the interposition cue, may not cast the proper (or even any) shadows on the real-world video objects (and vice versa), and the lighting model used for the graphics may not be consistent with the actual lighting at the remote site. In theory it is possible to compensate for these factors computationally, but in order to do so it is necessary, once again, to have an accurate model of the real objects situated at the remote site.

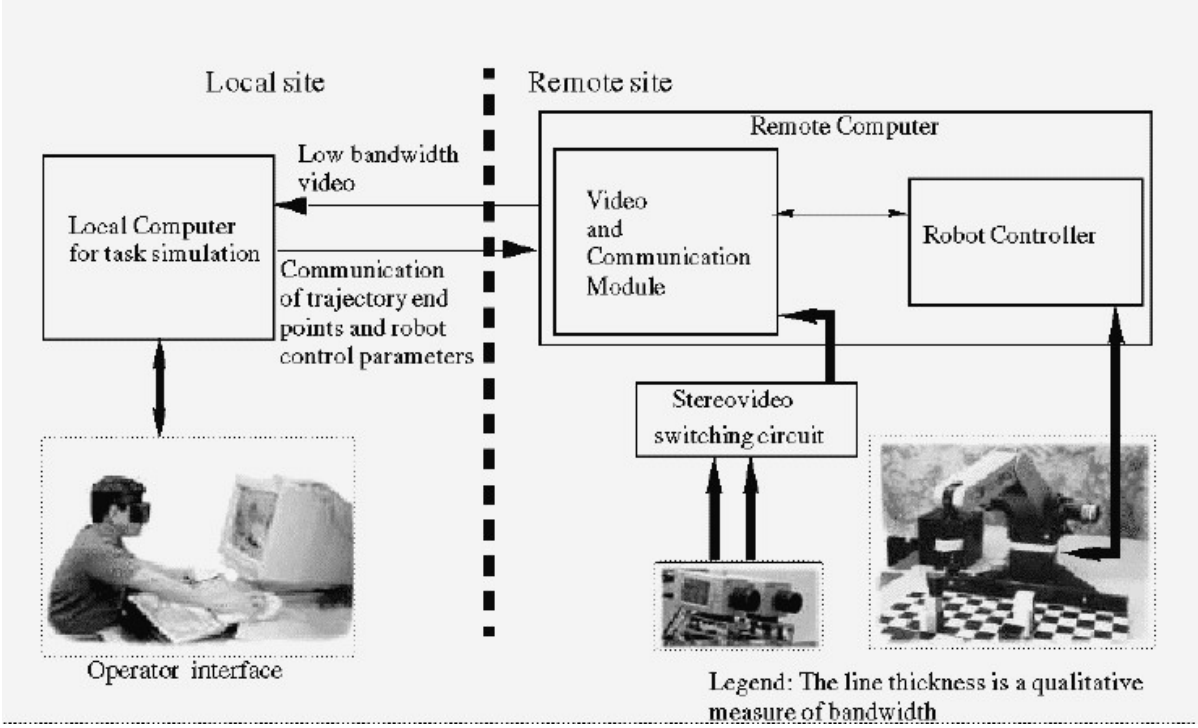

FIGURE 2. *Schematic diagram of ARTEMIS (**Augmented Reality TEleManipulation Interface System)***

In unstructured environments, therefore, the above mentioned depth cues are very hard to model and then replicate in graphics, to consistently match them with the video images. Even though other graphic depth cues, such as linear perspective and size constancy cues, can readily be made consistent with the video images, these monoscopic cues alone are not usually sufficient to adequately convey unambiguous depth information. Given the potential for monoscopic mismatches, stereoscopic displays become increasingly important, due to the contribution of additional binocular depth cues. When added to the monoscopic perspective and size familiarity cues, the binocular cues therefore improve the operator's ability to resolve some of the inconsistencies between the graphic and the real depth cues, and thus accurately to judge relative depths between graphical objects and video objects within a single AR display.

3.5. Interactive Modelling of Remote Site

The use of an AR based *virtual pointer* for interactive coordinate measurement at a remote site was originally reported by Drascic and Milgram5. Whenever the stereographic cursor (i.e. virtual pointer) is overlaid on a stereovideo image and interactively aligned by the operator with a point on an object in the SV image, the 3D coordinates of that point can be measured and recorded. This is possible because the perspective viewing parameters of the graphics are calibrated and registered with the video image, as described below. This allows the binocular parallax of the graphic cursor to be used to calculate the 3D coordinate values of associated points. These coordinate values can then be used for local programming and simulation of telerobot manœuvres, as described in Section 4. For example, if the values of any three coordinate points on a plane which is visible in SV are obtained interactively, these coordinate points can be used to create a quantitative model of that plane's position and orientation in 3D space. Such a model of the plane can then be used, for example, to prevent collisions during task simulation.

3.6. Calibration and Registration

At the remote site the stereovideo (SV) cameras should be configured not only to optimise the view of the site, but also to emulate the stereoscopic graphics (SG) through proper calibration and registration. In other words, before the SG images can be

generated, the graphic software must first be *calibrated* with the video cameras at the remote site, to ensure that the graphic images are drawn with the identical perspective projection parameters as those which determine the video camera images. These parameters consist of the camera field of view, the aspect ratio, and the origin of the video image coordinate system. Furthermore, since the graphical models are rendered in a separate graphical coordinate system, this must be matched with the world coordinate system of the remote site, such that all six degrees of freedom of the origins and the axes of both coordinate systems overlap on the resulting mixed display of video and graphics. This procedure is referred to as *image registration*. For example, in Fig 1. two images are shown in a single frame. One is a video image of a tabletop robot. The other is a *calibrated and registered* graphical image of the same robot, which has been overlaid on the video image in order that the graphical robot appears as if it is present at the same remote site.

The calibration and registration procedures together ensure that the base link of the SG robot model coincides exactly with the base link of the actual telerobot observed in the SV image. If the joint angles of the graphical model are set equal to those of the real robot, the entire SG image of the robot overlaps the SV image of the real robot as shown in the figure. By changing the joint angles of the kinematically modelled robot, its SG image can easily be manipulated relative to the underlaid SV image, while the base links of both robot images remain fixed with respect to each other. Although the graphic model of the robot and other objects can be rendered either with essentially transparent wireframe polygons or with filled polygons, the former is most often employed. For a detailed discussion of the perceptual issues which limit the use of filled polygons in such AR displays, see [Rastogi 1995] 10).

4. ARTEMIS: Telerobotic Control with Augmented Reality

4.1. Telemanipulation Through Simulation

In this section we describe the various components of ARTEMIS, with reference to the block diagram of the overall system given in Fig. 2. According to the present scheme the operator carries out a particular telerobotic task element by first simulating it graphically. This is done by manipulating the end effector of the stereographic (SG) robot such that the graphical robot arm appears to move within the background SV space. Whenever the end effector of the SG robot is moved to a new position, the real telerobot can easily be configured identically to the graphical one, as desired, on the basis of the SG robot's joint and end effector coordinates. To create an off-line programming segment, the operator takes the end effector of the SG robot to a succession of new locations within the task space and marks each of these as a trajectory endpoint. In order to manipulate a real object, the end effector of the graphical robot is aligned with the SV image of that object and the graphical gripper is commanded to close around it. The accuracy of the critical task of aligning the graphical gripper in depth depends primarily on the various depth cues discussed above. (For a discussion of the importance of such factors as the stereoscopic depth resolution of the display for determining the precision of object grasping, see [Rastogi 95][10]).

Since no prior model of objects in the task space is assumed, the operator is able to avoid collisions during off-line programming only on the basis of visual information directly perceivable in the display. The HO must therefore create the trajectory end points such that during the straight line motion between each starting point and end point of a trajectory segment there is no obstacle in the path. At the end of each off-line programming segment, the final sequence of trajectory endpoint commands defines a complete path, which can then be communicated for execution to the robot at the remote site. (The trajectory command sequence also incorporates additional parameters defining execution-delay and gripper state (i.e. fingers open or closed) for each trajectory end point.) One of the advantages of off-line programming is that, prior to execution, this sequence of trajectory commands can be "replayed" against the background of the current real SV image for verification, if required.

To summarise, the ARTEMIS approach serves to remove the human operator from prolonged closed loop manual control involvement by providing her with the means for communicating spatial instructions to the real robot in *non-real time.* Furthermore, unlike other telerobot programming schemes involving off-line graphical simulation, the ARTEMIS approach does not require extensive quantitative specification of the geometrical shapes and locations of target objects relative to which the robot must be manœuvred.

4.2. Example: A Pick-and-Place Task

In the following example a pick-and-place task in an unstructured environment has been selected, to illustrate how ARTEMIS can be used to carry out augmented telerobotic control. The reason for choosing a pick-and-place task is because it involves a variety of subtasks, including gross robot manipulation in 3D task space, precise alignment of the end effector with the object to be picked up, grasping of the object, robot manipulation after the object has been gripped in the fingers, and alignment of the picked object with another surface for subsequent placing. All segments of the example are accompanied by sample screen snapshots in

Fig. 3.

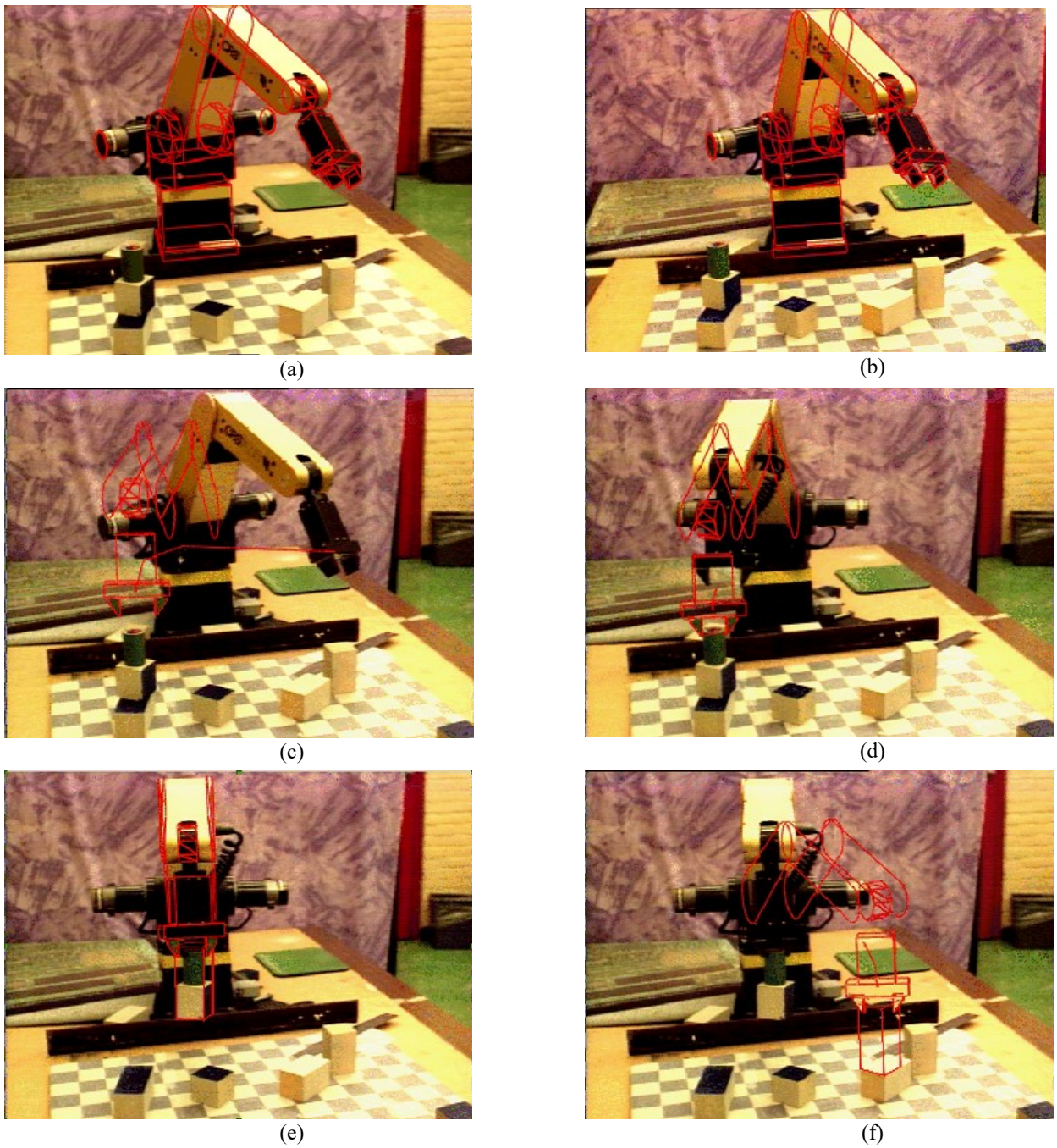

FIGURE 3. *An illustration of pick-and-place task simulation with augmented telerobotic control.*

- Fig. 3 shows the SG wireframe model of the robot overlaid on the SV image of the task space. Note that the base of the graphical robot coincides exactly with the video image of the real telerobot base. As an indication of the extent of binocular parallax used, both Fig. 3 (a) and (b) are presented here, representing respectively the left and right eye views of the composite AR image.
- To demonstrate the concept of off-line trajectory generation, Fig. 3(c) is presented (left stereoscopic view only), showing a simulated path depicted as a sequence of straight white lines joining trajectory end points (white circles). The path is generated by manipulating the SG model within the SV task space. This phase illustrates the gross manipulation subtask of bringing the robot to the vicinity of the object.
- Fig. 3(d) shows the real telerobot in the same position as the graphical robot at the end of its trajectory in figure Fig. 3(c), after that path has been sent for execution. Fig. 3(d) also shows the end effector of the graphical robot in a subsequent state,

aligned with the real video object to be picked up, with the graphical gripper closed. Following this step the end effector is raised, on the assumption that the object has been successfully "virtually grasped". Recall that the simulated path can again be verified for correctness by "replaying" the sequence before it is sent to the remote site for execution.

- Fig. 3(e) illustrates the system status after the object has successfully been picked up by the telerobot. In the figure a generic wireframe cube has been added adjacent to the end effector and positioned around the edges of the picked object. The objective of this step is to set the transformation matrix between the graphical gripper and the edges of the wireframe graphical cube. These edges of the wireframe cube are later used as an approximate model of the *picked object,* to perform the subsequent placing subtask. It is important to note that the wireframe cube is not an accurate model of the picked object, but contains only certain minimal features, such as the three perpendicular lines at one corner, to model the picked object partially.
- After partially modelling the picked object, the graphical robot is further manipulated to generate a path from the current position to a position in which the edges of the modelled wireframe cube are aligned with the surface. The graphical gripper is then opened and the graphical robot raised. Fig. 3(f) shows the wireframe cube being aligned with the surface on which the real object is about to be placed, while the real object in the meantime remains suspended at the endpoint of the previous trajectory sequence. Following simulation, the new set of instructions is sent to the robot for execution, resulting in the real object being successfully placed on the surface shown.

5. Advantages of ARTEMIS

The concept of teleoperation by means of stereoscopic augmented telerobotic control offers many advantages over direct manual control. Some of these are summarised as follows:

- *Display augmentation:* As a fundamental feature of ARTEMIS, the use of stereo displays improves the quality of presentation of depth and aids in identification of unknown objects, gradients and orientation of objects in the task space. The graphical simulation capability provides the opportunity for integrated display of information, by expressing a direct physical relationship between the simulated graphical robot model and real objects in the video display. The display can also be augmented by the presentation of additional task related information, such as distances and clearances in the remote site, working envelopes of the robot, preprogrammed trajectories, etc.

- *Reduced communication bandwidth:* One of the major problems with many teleoperation systems arises due to the absence of sufficient bandwidth between remote and local sites. In many cases this problem may preclude entirely the transmission of high bandwidth video images. With ARTEMIS, however, if the remote task space remains relatively unchanged within the duration of a particular trajectory sequence, the operator can perform graphical simulations locally, using static video images only. This is because the video images need to be updated only when instructions for a new configuration are sent to the telerobot, for the purpose of confirming commanded manœuvres and changes in task space. Since such updates can occur at a rate much slower than that required for direct manual control of the telerobot, the entire operation can be carried out essentially *independently* of system bandwidth.

- *Invariance to time delay:* Related to the above point, ARTEMIS can be used to provide the operator directly with control feedback from the graphical simulation rather than having to wait for feedback from the remote site. Under such circumstances the operator's manipulation performance should not be affected by either the magnitude, constancy or uncertainty of any time delays in the system.

- *Reduced manipulation errors:* One of the clear benefits of off-line task simulation is an expected reduction in programming errors, which might otherwise result in the robot's reaching an undesirable configuration or even colliding with some obstacle or potentially hazardous object. Such errors should be significantly reduced by making use of the playback of the graphical robot path within its real task space.

- *Higher levels of control.* With this scheme the operator is taken out of the direct spatio-temporal human-robot control loop and instead controls a graphical simulation of the robot at the local computer. Because this concept allows the HO in principle to send higher level of instructions, such as composite trajectory end points, to the remote computer, such a scheme can therefore accommodate higher levels of remote autonomy, where feasible. In such cases the operator essentially complements those autonomous functions by becoming a system monitor rather than just a controller.

6. Conclusion

ARTEMIS, a new augmented reality based interface for teleoperation in unstructured environments, has been described in this paper. This interface system combines both stereoscopic graphical modelled data and (unmodelled) stereoscopic video signals to enable off-line planning and simulation of teleoperation tasks, even in the absence of a world model of the task space.

7. Acknowledgments

This project has been supported by the Manufacturing Research Corporation of Ontario (MRCO), as well as by contract W7711-7-7009/01-SE with the Defence and Civil Institute of Environmental Medicine (DCIEM), Downsview, Ontario.

8. References

1. Boff K. R., Kaufman L., and Thomas J. P., [1986] "*Handbook of Perception and Human Performance" Volume1., John Wiley and Sons.*

2. Draper J. V., Handel S., and Hood C.C. [1991] "Three Experiments with Stereoscopic Television: When it Works and Why", *Proceedings of IEEE International Conference on Systems, Man and Cybernetics, Charlottesville, Virginia*

3. Drascic D. [1991] "Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing" *Proceedings Human Factors Society 35th Annual Meeting, 1367-1371.*

4. Drascic, J. Grodski, P. Milgram, K. Ruffo, P. Wong and S. Zhai, "ARGOS: A display system for augmenting reality*", ACM SIGGRAPH Tech Video Review, Vol. 88: InterCHI '93 Conf. on Human Factors in Computing Systems, Amsterdam, April 1993.*

5. Drascic D., and Milgram P., [1991a] "Positioning accuracy of a virtual stereographic pointer in a real stereoscopic video world", *SPIE Volume 1457 Stereoscopic Displays and Applications II, 302-312.*

6. Milgram P., Zhai S., Drascic D., and Grodski J. J., "Applications of augmented reality for human-robot communication", *Proc. IROS'93: International Conf. on Intelligent Robots and Systems, Yokohama, Japan, 1467-72, 1993.*

7. Milgram, P., Takemura, H., Utsumi, A., Kishino, F., "Augmented Reality: A class of displays on the reality-virtuality continuum", *Proc. SPIE Vol. 2351-34, Tele-manipulator and Telepresence Technologies, 282-292, Boston, Oct. 1994*

8. Milgram. P., Drascic D., Grodski J. J., Rastogi A., Zhai S., and Zhou C., "Merging real and virtual worlds", *Proc. Imagina'95, Monte Carlo, 221-230, 1995.*

9. Rastogi A., Milgram P., Drascic D., and Grodski J. J. [1993] "Virtual Telerobotic Control" *Proceedings of the Knowledge-Based Systems & Robotics Workshop, Nov. 14-17, 1993; 261-269*

10. Rastogi A. [1995] "Design of an interface for teleoperation in unstructured environments using augmented reality", *M.A.Sc. Thesis, Department of Industrial Engineering, University of Toronto, Toronto.*

Figure 3 (a) & (b) are left and right eye views of the stereoscopic display, and Figure 3 (c), (d), (e) and (f) are only left eye views of the stereoscopic display. Refer to accompanying text for explanation.