*41st Annual Meeting of Human Factors and Ergonomics Society, Albuquerque, New Mexico. Sept. 1997.*

# PERCEPTUAL EFFECTS IN ALIGNING VIRTUAL AND REAL OBJECTS IN AUGMENTED REALITY DISPLAYS

Paul Milgram and David Drascic
Department of Mechanical and Industrial Engineering
University of Toronto
Toronto, Ontario, Canada

The concept of Augmented Reality (AR) displays is defined, in relation to the amount of real (unmodelled) and virtual (modelled) data presented in an image, as those displays in which real images, such as video, are enhanced with computer generated graphics. For the important class of *stereoscopic* AR displays, several factors may cause potential perceptual ambiguities, however, which manifest themselves in terms of decreased accuracy and precision whenever virtual objects must be aligned with real ones. A review is given of research conducted to assess both the magnitude of these perceptual effects and the effectiveness of a computer assisted *Virtual Tape Measure (VTM)*, which has been developed for performing quantitative 3D measurements on real-world stereo images.

## BACKGROUND

This paper deals with visual perceptual factors which influence performance when using Augmented Reality (AR) displays as a remote measurement or control tool in application domains such as telerobotics and medicine. AR displays are defined here as a subset of the class of "Mixed Reality" (MR) displays, which in turn are defined within the larger context of the Reality-Virtuality (RV) continuum (Milgram & Kishino, 1994). As depicted in Fig. 1, the RV continuum is presented as a framework for describing the spectrum of cases that define whether the primary world being experienced by an observer is *real* or *virtual*. One way to display real world objects is by scanning, transmitting and reproducing image data, as is the case with ordinary video displays[1] -- without the need for the display system to "know" anything about the objects. Another way is by viewing real-world scenes either directly or via some optical lens system. Virtual images, on the other hand, can be produced only if the computer display system generating the images has a *model* of the objects being portrayed.

Fig. 1 shows that MR refers to the class of all displays in which there is some kind of combination of real and virtual environments. Within this context, the meaning of the term *"Augmented Reality"*, depicted on the left side of the continuum, becomes quite clear: AR displays are those in which the primary image is of a real environment, which is enhanced, or augmented, with computer-generated imagery. As shown in the figure, in other words, the difference between the purely real environment on the left, depicting a video image of a person next to a robot, and the AR example to the right is the addition of the graphical robot on the table. In general, Augmented Reality enables one to make virtual images appear before the viewer in well specified locations in the real world image. Such images can display task related data, or can serve as interactive tools for measuring or controlling the environment, using either direct viewing (DV) or head-mounted video "see-through" displays or ordinary display monitors.

In contrast to AR, *"Augmented Virtuality" (AV)* displays are those in which a primarily virtual environment is enhanced, or augmented, through some addition of real world images or sensations. Such additions can take the form of directly viewed (DV) objects, where users might see their own

[1] Note that, although we are limiting our discussion here to *visual* displays, similar classfications may be made with respect to other display modalities. For example, *real* sound sources may be directly transduced or replayed, whereas a *virtual* sound source could be produced through computer modelling and synthesis.

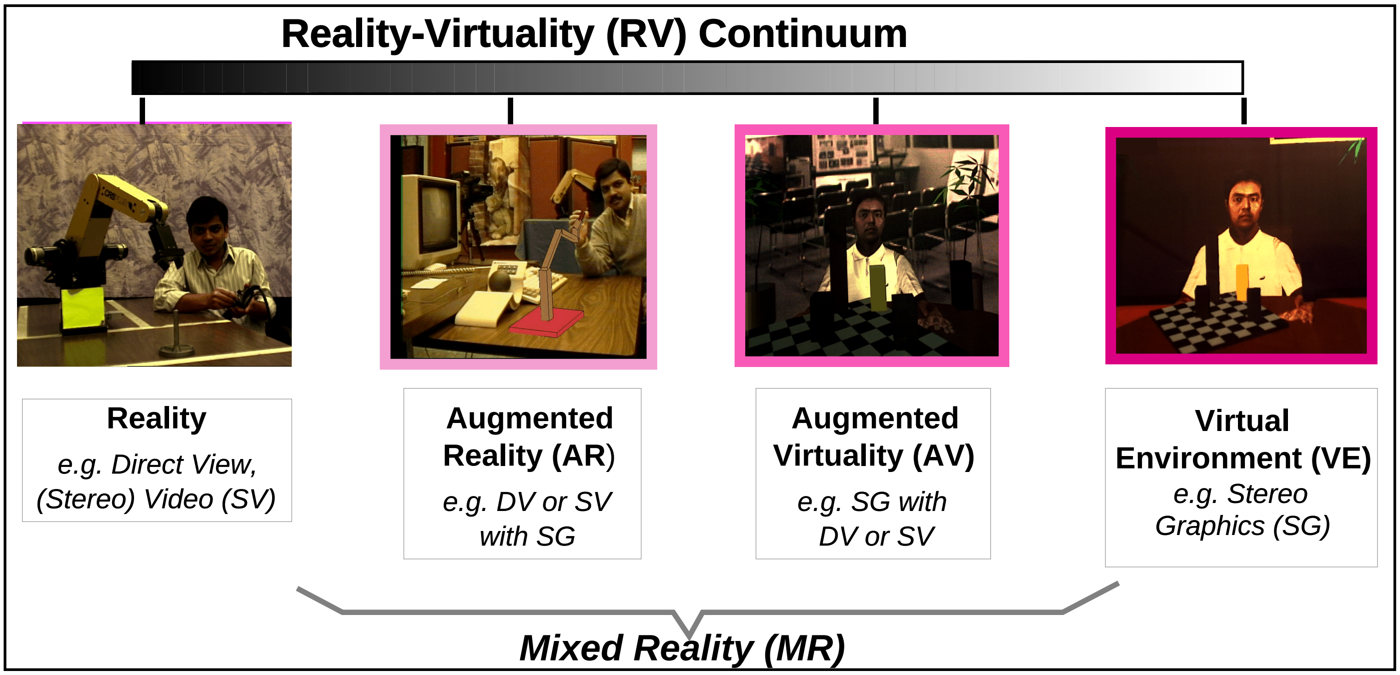


Figure 1: Simplified representation of the Reality-Virtuality (RV) Continuum, showing how real and virtual worlds can be combined in various proportions, according to the demands of different tasks.

limb instead of a computer-generated simulation, as is common with surround type virtual environments (VE's) where one might reach into the scene to grasp an object with one's own hand. Another AV mode is when video images are added to otherwise completely simulated displays. This concept is shown in Fig. 1 by the completely virtual (modelled) image at the extreme right side of the RV continuum, which is augmented by adding an (unmodelled) video background in the AV example to the left.

In this paper we deal with (visual) Augmented Reality displays only, and we further limit ourselves to the special, but very significant, case in which all viewing systems are *stereoscopic*. Our particular interest lies in situations in which the available 3D cues do not completely support each other, and may even be in conflict, thereby leading to distorted perceptions of depth, distance or shape. (Drascic & Milgram, 1996).

## REAL-VIRTUAL ALIGNMENT ERRORS IN AUGMENTED REALITY

One class of tasks which is particularly influenced by such distortions is that of aligning *virtual* objects with *real* ones (RV alignment). In AR environments one may require this capaibility for visualising how, as shown in the AR example of Figure 1, a virtual 3D graphic object would appear against the real 3D video (SV) background into which the model has been constructed to fit. In a conceptually similar application, we have superimposed simulated human operator mannequins onto real SV workplaces, for the purposes of ergonomic workplace analysis. In such cases the important perceptual issues involve having the virtual mannequin appear to fit in properly with the background and having its limbs appear to make contact realistically with the floor, chairs, tools and other instruments.

In other cases, it may be necessary to make reliable 3D measurements of the dimensions or locations of various objects within the SV image, as well as distances between those objects. This latter capability comprises the essence of our AR *Virtual Tape Measure (VTM)* (Milgram et al, 1997), one of the fundamental capabilities of our ARGOS (Augmented Reality through Graphic Overlays on Stereo-video) display system (Drascic et al, 1993). One important application of the VTM, presented elsewhere in this proceedings

(Kim et al, 1997), is for intraoperative measurement of anatomical structures during minimally invasive surgery. Yet another extension of the VTM concept is the ability to simulate a complete overlaid virtual remote robot, in order to carry out off-line teleprogramming over low bandwidth communication lines (Rastogi et al, 1996).

In a separate study (Drascic & Milgram, 1996), we have proposed an exhaustive classification of pertinent perceptual issues affecting virtual-real alignment performance in MR displays involving stereoscopic video (SV), stereoscopic graphics (SG), and direct view (DV), using head-mounted displays (HMD's), desktop monitors and large screen projection systems. In summary, these issues are classified according to:

- Implementation Errors: These errors comprise perceptual inaccuracies due to calibration errors, calibration mismatches, and interpupillary distance errors.
- Technological Limitations: These errors comprise static and dynamic registration mismatches, restricted fields of view, limitations and mismatches of resolution and image clarity, luminance limitations and mismatches, contrast mismatches, size and distance mismatches, depth resolution limitations, vertical alignment mismatches and viewpoint dependency mismatches.
- "Hard" problems: These include object interposition failures, expanded depth of field, absence of accommodation, accommodation-vergence conflicts, accommodation mismatches and absence of shadow cues.

## EXPERIMENTAL INVESTIGATIONS

Due to the criticality of the real-virtual object alignment issue, we present the results of a set of empirical investigations of the precision and accuracy of RV alignment in an AR environment. Two classes of experiments were performed: one to compare the precision and accuracy of RV alignment using human visual perception alone, and the other to assess the effectiveness of machine aided versus unaided RV alignment.

### *Unassisted RV Alignment Performance*

The initial experiment addressed the pointing accuracy of a virtual SG pointer with respect to real SV images in a depleted environment -- that is, one in which all cues but binocular disparity were removed (Drascic & Milgram, 1991). The experiment was a method of adjustment task involving the aligning of two vertically oriented pointers. A 2x2 experimental design was used, comprising a combination of real and virtual pointers {RP, VP} and real and virtual targets {RT, VT}. The main conclusions reached in that experiment were, in terms of mean error, that is, pointing *accuracy*, that there were no significant differences among the four conditions. However, there was a small but consistent mean error, which implies that subjects are somewhat inclined towards placing the pointer in front of the target (i.e. closer to themselves). The magnitude of that bias was only approximately 20 arc-seconds, however, which, in terms of screen units in that experiment, corresponded to an error of about 1/7 of a pixel.

With respect to standard deviation, that is, in terms of pointing *precision*, the only significant effect appeared to be not perceptual but due to the different interfaces used for controlling the virtual pointer (VP) in one set of cases and the real pointer (RP) in the other. The important overall conclusion from that first experiment was that the unaided subjects were in fact able to align virtual pointers with real targets essentially just as well as they were able to align real pointers with real targets, using visual perception alone.

### *Assisted RV Alignment Performance*

As promising as the original results were, two weaknesses are apparent:

a) The measurements performed in the depleted visual environment of the experiment, where only binocular disparity cues and, to a lesser extent, size cues were present, are not necessarily representative of measurements in actual real-world SV scenes.

b) Even if an operator is in principle *capable* of performing well using the Virtual Tape Measure (VTM), s/he may not do so *consistently.* Furthermore, as long as the computer is not provided any data about the real world, it has no way of checking on the operator's performance and thus ensuring an acceptable level of reliable performance during actual operations.

It is for these reasons that we have developed a *computer-assisted* version of the VTM (Milgram et al, 1997). The assisted version of the VTM is based on interactive invocation of a set of computational vision tools, which allow the HO to request that the computer provide an alternative version of the actual 3D location of the virtual SG pointer relative to a designated real SV object. The

HO is then free to accept the machine version, remain with her own original perceptual estimate or, ideally, to confirm agreement of the two estimates.

*Precision + Accuracy Assessment Experiment*

An experiment was carried out to evaluate both assisted and unassisted modes of the VTM, using real world targets under representative (that is, not ideal) conditions of lighting, camera alignment and target contrast ratio. All subjects (N=5) were all experienced with the Virtual Tape Measure. Measured target separations ranged across small (10°) to moderate (20°) distances relative to the camera system's optical centroid. In addition, measurements were made not only near the convergence point of the stereo camera system, but also in front of it (crossed disparity) and behind it (uncrossed disparity). All measurements were repeated using both the assisted and the unassisted VTM.

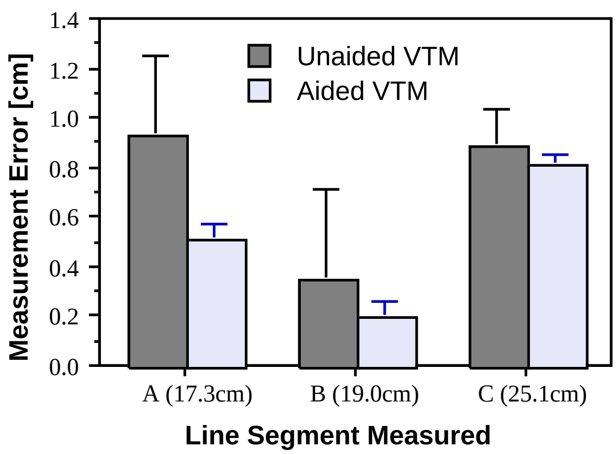


**Figure 2**: Magnitude of Measurement Errors for the different line segments, along with standard error bars. Line segments A and C each had large depth components. Line segment B did not.

As with earlier experiments, performance with the VTM was assessed in terms of both accuracy (i.e. central tendency) and precision (i.e. dispersion). The results for the former are presented in Fig. 2. The most important result of that experiment is that a significant measurement bias was again detected, for both aided and unaided VTM's ($F(1,4)=28.9$, $p=.006$). No significant differences were found between the magnitudes of this bias for unaided and aided VTM's, however. It is important to note that the mean magnitude of the bias was +0.62 cm, which translates in this case to a mean overestimation error of 8%. As expected, no significant differences were found with respect to the type and magnitude of the actual distances measured (ranging between 17 and 25 cm).

In terms of precision of VTM placement, an analysis was done on the log standard deviation estimates acquired from the error measurements, as shown in Figure 3. The most important result obtained from that analysis is that there was a significant difference between the aided and unaided virtual tape measurement standard deviations ($F(1,2)=47.3$, $p=.02$). The other noteworthy result from the Fig. 3 analysis is that there was also a significant difference ($F(2,4)=15.1$, $p=.014$) due to the type of measurement made. This difference was essentially due to whether or not measurements were made in the same frontal plane (segment B in the figures) or in the depth direction (segments A and C).

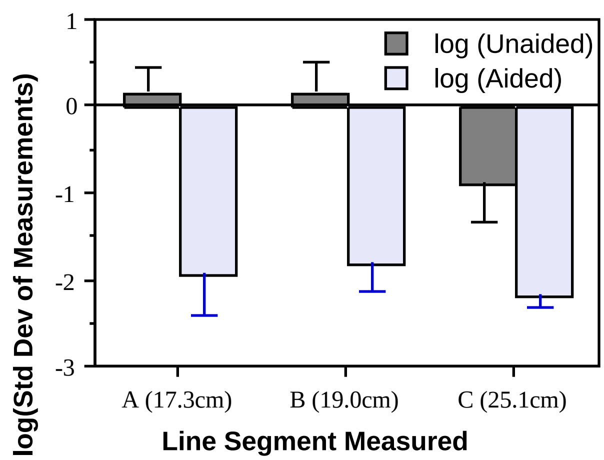


**Figure 3**: Log of the standard deviations of the measurements, along with standard error bars for the grouped deviations.

## CONCLUSION AND DISCUSSION

The general conclusion to be drawn from our research to date is that performance with the Virtual Tape Measure is generally acceptable, with a few exceptions. First of all, the spread of data with the aided cursor is significantly less than that without computer aiding, indicating that performance with the aided cursor is 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 small but consistent positive error; i.e. an overestimation of the measured distances. Although one interpretation of this result is that there is a potential systematic error in our stereo camera calibration system, it is our belief that this is rather most likely 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.

In spite of the bias which was detected, it is important to take note of the actual error magnitude plots relative to the distances measured, which for many represent the most significant results at a practical level. 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.* Significant improvements are to be expected, however, if major changes were to be made to the camera alignment parameters and the focal lengths used, and if optical distortions were taken into account.

As the technology for implementing Augmented Reality becomes more accessible and new areas of application are demonstrated, the use of AR displays is expected to continue to accelerate (Barfield et al, 1995). In addition to the robotics applications indicated, another important practical domain is in medicine, especially computer-aided surgery. In our own lab, for example, we are currently testing the feasibility of using the AR Virtual Tape Measure for intraoperative measurements during minimally invasive micro-neurosurgery (Kim et al, 1997). In other labs, efforts are underway to provide AR overlays of preoperatively imaged brain data during neurosurgery, or computer generated planning models during cranial reconstruction surgery. In all cases, one of the critical parameters which will determine the acceptance of this technology by practitioners is whether or not it will be feasible to make computer generated virtual objects appear alongside real ones and, as required, in alignment with them. As outlined here, the various factors which influence this perception form a critical area of research.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the generous support provided for this research by the Dr. Julius Grodski and Defence and Civil Institute of Environmental Medicine (DCIEM) , the Institute of Robotics and Intelligent Systems (IRIS) and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

## REFERENCES

1. Barfield, W., Rosenberg, C., Lotens, W.A., "Augmented-Reality Displays", Chapter 14 in W. Barfield and T.A. Furness III (ed's), *Virtual Environments and Advanced Interface Design,* Oxford Press: NY, 1995.
2. Drascic, D., Grodski, J., Milgram, P., Ruffo, K., Wong, P. & Zhai, S. "ARGOS: A display system for augmenting reality", *ACM SIGGRAPH Technical Video Review, Vol. 88(7), InterCHI'93 Conf. on Human Factors in Computing Systems,* Amsterdam, 1993.
3. Drascic, D. & Milgram, P. "Positioning accuracy of a virtual stereographic pointer in a real stereoscopic video world", SPIE Vol. 1457: Stereoscopic Displays and Applications II, 302-312, 1991.
4. Drascic, D. & Milgram, P. "Perceptual issues in augmented reality", *SPIE Vol. 2653: Stereoscopic Displays and Applications VII and Virtual Reality Systems III,* 1996.
5. Kim, M., Milgram, P. & Drake, J.D. "Computer assisted 3D measurements for microsurgery", Proc. 41st Annual Meeting of Human Factors and Ergonomics Society, Albuquerque, NM, Sept. 1997.
6. Milgram, P. & Kishino, F. "A taxonomy of mixed reality visual displays", *IEICE Transactions on Networked Reality,* E77-D(12), 1321-1329, 1994.
7. Milgram, P., Yin, S. & Grodski, J.J. "An augmented reality based teleoperation interface for unstructured environments", *Proc. American Nuclear Society 7th Topical Meeting on Robotics and Remote Systems,* Augusta, Georgia, 1997.
8. Rastogi, A., Milgram, P. & Drascic, D. "Telerobotic control with stereoscopic augmented reality", *SPIE Vol. 2653: Stereoscopic Displays and Applications VII and Virtual Reality Systems III,* 115-122, 1996.