

Target Selection in Augmented Reality Worlds

Jamie Sands, Shaun W. Lawson, and David Benyon

Human Computer Interaction Group,
School of Computing, Napier University,
10 Colinton Road, Edinburgh,
United Kingdom, EH10 5DT.
{j.sands,s.lawson,d.benyon}@napier.ac.uk

Abstract. This paper describes our work in developing the human-computer interface of an augmented reality (AR) system that has been designed to facilitate the accurate selection of real world locations in 3D space using 3D cursors. We detail the results of an experiment undertaken to highlight the benefit of certain features integrated into the design of a 3D cursor and the interface employed to manipulate the cursor. We conclude that object selection is aided by adding additional synthetic depth cues to the cursor and also by using constraints on the number of degrees of freedom exhibited by the cursor.

1 Introduction

Augmented reality (AR) is the term used when fusing real world environments with context-aware computer generated information. The underlying idea is that a human's task, or experience is augmented in some way by the use of overlaid computer graphics, text or even audio. The potential applications of AR are numerous and wide ranging and readers are directed to a review of recent advances in the field [1] for a summary of these. Despite the wealth of potential applications, AR is still a relatively infant technology with many issues still poorly understood. From an engineering point of view, for example, it can be said that the ultimate goal of AR research is to attain perfect dynamic registration between the real and virtual worlds. This is indeed the focus of much current research, including some of our own [2]. However, the problems of registration have tended to over-shadow many other important issues including the physical, physiological and psychological (cognitive) limitations of using AR systems. Indeed, detailed studies of Human Computer Interaction (HCI) with AR systems have appeared infrequently in the literature. Above all, the issue of usability of AR systems and the associated problem of applying augmented reality to a task to do *work*, has received very little attention.

Whilst common problems between AR and Virtual Environments/Reality (VE/VR) exist (selection of input/output devices, choice of tools and metaphors for instance) we believe that AR presents an entirely new set of problems to the HCI community. These problems stem from the fact that an AR display contains, by definition, an amount of real world content and, in many cases, very little prior information is known about this content. This novel setting, unlike VR, combines an environment and objects that conform to the physical laws of the real world and virtual overlaid



Fig. 1. Panoramic picture of the surface of Mars, landscapes like this could be remotely mapped from Earth using 3D AR cursors.

graphics that may not. The accurate manipulation of virtual objects with respect to the real world therefore represents a problem not currently addressed by VR work.

Consider a user interacting with a visual display of a VE - an interface designer can make use of the fact that the graphics system can detect, via the underlying projection model employed, when a user positions a selection tool over the top of an object. This allows the designer to highlight the object to let the user know that it has been selected. Additionally, the selection tool in a VR system can become occluded by or will occlude objects in the environment and is subject to the same rendering rules (shadowing, shading and coloring) as all other objects within the environment. When selecting a real object in an AR display none of the above factors hold true: a selection tool will simply occlude all objects that lie 'behind it' no matter their relative positions in 3D space. The selection tool is also rendered according to an approximation, at best, of the real world lighting conditions. The end result is an interface that conveys poorly constructed, or even conflicting, depth cues to the user. Couple these problems with the already well known issues and deficiencies of typical interface devices such as Head Mounted Displays (HMD's) and 3D mice then this results in an interactive system that is potentially very difficult to use.

Despite these inherent problems research into AR systems is a worthwhile pursuit. In comparison to VR, AR has an almost limitless scope for application. Whereas Brooks suggests that VR has only two main domains of application, training/simulation and the presentation/demonstration of environments or objects [3], AR applications include anything where an advantage is gained by combining overlaid graphics and the real world. In his 1997 and 2001 surveys of AR Azuma presents a good synopsis of many areas of proposed applications [1,4].

An essential element of successful AR systems is the ability for the real and virtual to appear to co-exist in the same space. An area of particular interest is the accurate alignment of 3D cursors with the real world in order to measure distances. One particular application that could utilize the accurate positioning of augmented 3D cursors is the mapping of remote inaccessible or inhospitable environments such as the surface of Mars, Figure 1. 3D AR systems offer a repeatable method of accurately examining such environments without the need for the presence of a human user and can be conducted 'offline'. Users can measure and map these environments from a remote location using only the received stereo images, thus eliminating the need for complex and expensive mapping technology.

In this paper we describe our efforts to produce a set of general guidelines, independent of problems associated with system variables such as display devices, to aid AR researchers in the development of selection tools such as virtual cursors. In

particular we describe recent work in addressing the accuracy of object selection via the addition of artificial depth cues in a stereoscopic display. Our experiment was designed to assess some aspects of the system chosen to manipulate our virtual cursor. We conclude by stating the findings of the work, which have confirmed a number of well-known HCI principles. We also describe our current efforts to move our users away from the desktop and into a walking, wearable AR environment.

2 The Importance of Object Selection in AR

Using virtual tools to select real objects is a task common to numerous proposed AR applications. The accurate selection of chosen targets precludes the ability to label objects, people and buildings, and the measurement of sizes and distances. Much of our own work has concentrated on the development of AR tools for the exploration of remote environments using tele-operated mobile vehicles [2,5,6]. These tasks require a human operator to select points in 3D space to a very high precision. However, given that the mobile vehicle can be accurately tracked, the technique is repeatable and could, potentially, be scaled to allow full mapping and characterization of complete remote scenes – the inside of buildings or planetary landscapes for instance.

Similar work to our own has been undertaken at the University of Toronto [7] with the primary objective being to determine whether subjects can accurately mark the location of real points in an unprepared space. This work, like our own features either a stereoscopic desktop display or HMD coupled with a 3D (often 6 degrees of freedom (DoF)) desktop mouse. Experiments in this area are notoriously difficult to repeat across research groups since numerous systematic sources of error are inherently present in any AR system. The sources of these errors include the calibration strategy employed, the tracking of any devices used, and distortions and flaws in the chosen display system.

Another related piece of work is the *Tinnith-Metro* system [8] under development at the University of South Australia. The developers of *Tinnith-Metro* have incorporated a glove-based interface into a walking AR system so that new virtual objects can be created. This can allow a subject either to completely model the existing world, or to create new augmented parts to real worlds. This system allows the user to walk about the environment whilst interacting with it at a very involved level.

2.1 Borrowing Object Selection Tools from VES

It is clear from the above discussion that the selection of objects will comprise an important set of interaction techniques in a wide range of future AR applications. However – how best to implement such interactions is, at present, unclear. The manual selection of objects in a 3D *virtual* environment (VE) is a well-studied HCI problem with many solutions having been proposed in the literature. Poupyrev et al, for instance, give taxonomy of object selection techniques in [9]. They classify relevant techniques as using either virtual hand or virtual pointer metaphors. Many such methods can work in VEs because the entire environment is known, whereas in AR the real element of the environment is often not known (we use the term *unprepared*). Therefore virtual pointer methods, such as ray casting for instance, are

inappropriate as they rely on knowing when a ray intersects a known object. An equivalent ray in a mixed reality environment would simply sit on top of all objects within the display.

Our 3D cursor is manipulated by the user until its “hot-spot” (for instance the tip of a cone) appears, to the subject, to coincide with the position of an object. In pure VE’s, 3D cursors are simply another virtual object in the environment – they are subject to the same collision detection, occlusion, shadowing and rendering rules as all other objects. In an unprepared AR environment however a virtual 3D cursor is subject to none of these rules – a cursor will occlude all real objects in the world no matter of their relative position. The end result of this is that users have difficulty in positioning a 3D cursor in an AR scene – largely, we believe, because the depth information available to the user exhibits conflicting cues. This remains true even when stereoscopic displays are used to present the mixed world to the user. Other factors such as the general awkwardness of 3D stereoscopic displays and interface devices such as 3D mice also serve to make cursor manipulation a difficult task.

3 Experimental Conditions and Procedure

In order to determine the effect of a number of variables in a typical object selection task using a 3D cursor we have recently conducted a series of experiments using our stereoscopic video see-through AR system [2]. Our experiments were designed to assess the following:

- The accuracy of subjects’ ability to position a 3D cursor in an unprepared mixed reality environment with no other extra information present to the user
- Whether the use of a synthetic shadow cast by the virtual cursor improved subjects accuracy in positioning of the cursor
- Whether the overlay of textual information regarding the cursors whereabouts in the world improved subjects accuracy in positioning of the cursor
- Whether adding user selectable constraints to the number of degrees of freedom (DoF) available when manipulating the cursor improved subjects’ accuracy in positioning of the cursor
- Whether the presence of an object of known size within the display improved subjects’ accuracy in positioning of the cursor
- Whether there were any correlations between subjects’ spatial ability (measured using a general purpose mental object rotation test) and cursor positioning accuracy.

An early version of our AR system has been described in detail in [2]. Our most recent system features two fixed color CCD cameras in a parallel configuration gazing at a real scene containing a number of targets in known real world locations. The left and right analogue PAL camera signals are overlaid with corresponding graphics streams (using analogue luma-keying devices) and presented to a user in a remote room via a stereoscopic HMD (a 5dt device). Usually, the only overlaid computer generated objects are a 3D cursor and textual information shown in a fixed location in the display. A static calibration of the system is performed as described in [2].

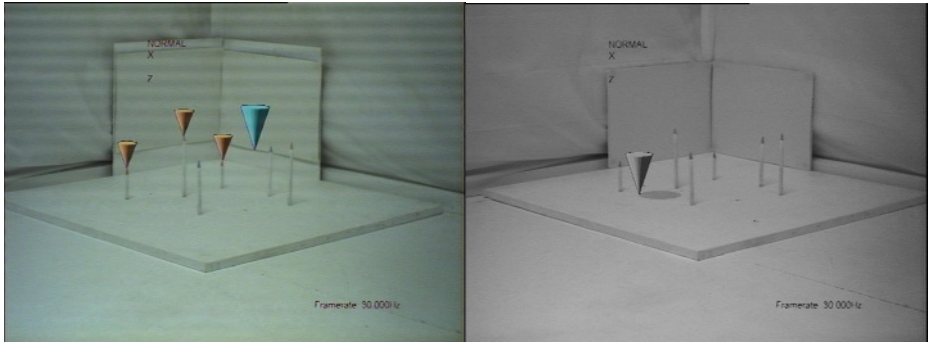


Fig. 2. Screenshots of base condition display containing 3 target cones and the shadow condition display.

Due to the luma-keying method employed in the overlaying of virtual graphics a blank, white scene, lit using diffuse lighting was employed to avoid areas of conflicting shadow and varying degrees of brightness. The blank scene also kept the depth information available to the user at a minimum.

The purpose of our experiments was to investigate subjects' accuracy in aligning a 3D virtual cursor with a number of real objects (targets) in a stereoscopic display. The targets consisted of the tips of a number of upturned sharpened dowels mounted on a rigid base (see Figure 2). Each individual experiment required the subject to drop a virtual marker at six such targets using a Spacemouse device to manipulate the cursor. A Spacemouse interface was chosen as it naturally offered 3DoF to subjects in a familiar desktop setting; the experimenters also had previous experience programming Spacemouse interface devices. Subjects' errors in the alignment tasks, in millimeters, were measured during eight experimental conditions using a within subjects repeated measures procedure. The eight conditions are described in Table 1.

The experimental conditions described in Table 1 aim to uncover which additional synthetic depth cue provides the user with the most useful extra information. These particular cues were chosen as each could be added to the 3D cursor without affecting the cursors movement or the users view of the cursors hotspot. They were also chosen as they did not appear to be application specific and therefore could be easily applied in any AR system requiring the accurate positioning of a cursor. The cues can be separated into two broad groups: cues that assist positioning by indicating a ground intercept and those that assist the user by offering scalar information. Figure 2 illustrates the correct shadow condition showing the indication of the 'ground' level.

Fifty-six subjects (thirty six male and twenty female) aged between 19 and 54 years took part in the experiment. No prior experience was necessary. No subjects reported blindness in either eye although visual acuity varied between subjects.

At the start of each experiment each subject was briefed on the task they were about to perform, shown how to use the Spacemouse interface and allowed to practice moving the cursor. Subjects were made aware that the tasks would not be timed. Once subjects confirmed they understood the experiment and could satisfactorily move the cursor they were instructed on how to use the stereoscopic HMD. Subjects were given the choice of having the text displayed to their left or right eye, and were also given the choice to wear or remove their spectacles as they saw fit. It was also

Table 1. List of experimental conditions

Condition Name	Description
Base condition	In this condition the alignment task was undertaken using a simple upturned cone cursor as shown in Figure 1.
Non-constrained 3D motion	In all other conditions including the base condition, subjects were able to constrain the motion of the cursor by switching off motion in x, y and z directions. However in this condition, the cursor moves freely in all DoF.
Co-ordinates display	In this condition a world centered (not user centered) text overlay indicated the absolute cursor position to the user in the upper right corner of the display.
Correct Shadow	This condition incorporated a ‘token’ virtual shadow cast by the cone cursor onto the experimental tabletop plane. The shadow position conformed to the real light source that was present in the real world (to the upper left of the subject’s viewpoint). The shadow moved and behaved, as would a real shadow except that it could not, of course, be occluded by real world objects. If the subject positioned the cursor under the surface of the table (at a position of $y \leq 0$), then the shadow disappeared.
Crosshairs Cursor	The Crosshairs condition replaced the upturned cone of the base condition with a 3D crosshairs with a central dot indicating the ‘hot spot’. The spars of the crosshairs were created along the lines of the x, y and z dimension axis.
Incorrect Shadow	Same as the Correct Shadow condition except that the shadow appeared in the ‘wrong’ location considering the position of real light source i.e. it appeared on the opposite side.
Scale	In this condition an additional object – a scale of known size - was added along the perceived x and z-axis. The scale comprised a series of black and white bands each 1cm apart.
Textured cursor	In this condition the cone cursor was enhanced by applying a texture to its surface, thus adding texture gradient cues.

stated that subjects could re-adjust the focus and position of the HMD as often as they wanted and at any time during the experiment sessions.

Each subject was then asked to perform the base condition; this condition was always performed first for a number of reasons. The base condition served to further familiarize the subject with the task prior to the more complex experimental conditions. As the base condition was conducted as a baseline measurement, against which the other conditions would be compared, this condition had to be devoid of any other information cues. Performing this condition first meant that no residual information could be carried over from the tasks containing extra cues. After completing the base condition subjects completed the six experimental conditions in a randomized order. The final condition was always the non-constrained 3D-motion condition; this allowed subjects the maximum possible practice with moving the

cursor in 3D before conducting this condition. Overall the time taken to complete instruction and experimentation was around 40 minutes with each task taking around 4 minutes each.

4 Experimental Results and Analysis

Table 2 shows the mean errors, standard deviations and variance for each of our experimental conditions. One of our primary objectives was to investigate the effect of additional depth information on alignment accuracy -therefore our analysis compares subject's performance in each condition with the base condition where no extra depth information was presented. Therefore the benefit of a single depth cue can be examined. As can be seen from Table 2 only the non-constrained 3D-motion condition has a higher mean, or performed worse, than the base condition, all the additional stimulus conditions performed better than base condition. Inspection of the standard deviations also shows that all conditions other than the 3D-movement condition had relatively similar degrees of deviation from the mean.

Table 2. Summary of results.

Condition Name	Mean Error /mm	Std./mm	Variance /mm
Base condition	12.41	8.47	71.65
Un-constrained 3D motion	46.82	34.65	1200.89
Co-ordinates display	8.38	6.74	45.37
Correct Shadow	6.97	6.09	37.13
Crosshairs Cursor	12.29	8.06	64.90
Incorrect Shadow	8.46	9.63	92.77
Scale	10.44	8.83	77.93
Textured cursor	9.25	7.05	49.66

A within subjects repeated measures analysis of variance (ANOVA) was conducted to determine any significant differences between the experimental conditions. Due to the highly varied nature of the data obtained and performance of subjects it was decided to remove all the outlying and extreme values. During testing, observations showed some subjects were unable to complete all the tasks in a similar fashion. As these discrepancies generally did not indicate an inability to complete the tasks correctly but instead, a loss of concentration or a misunderstanding it was decided that these results were not indicative of the overall performance. These values were replaced by the mean value, in total 37 results were replaced over all conditions. Removal of these values did not affect the levels of significance of any of the results. Statistical analysis found a significant effect of stimulus type (i.e. between conditions): $F(7,55) = 51.224$, $p < 0.001$. It was concluded that there exists a significant difference between the means of the experimental conditions.

Table 3. Significance results for the between condition paired t-tests.

Condition paired with Base Condition	Difference in Mean Errors /mm	Significance result for paired test
Non-constrained 3D motion	34.41	$t(55) = 7.38, p < 0.007$
Co-ordinates display	-4.04	$t(55) = -3.39, p < 0.007$
Correct Shadow	-5.45	$t(55) = -4.55, p < 0.007$
Crosshairs Cursor	-.124	$t(55) = -0.09, p > 0.007$
Incorrect Shadow	-3.96	$t(55) = -2.43, p > 0.007$
Scale	-1.98	$t(55) = -1.33, p > 0.007$
Textured cursor	-3.17	$t(55) = -2.51, p > 0.007$

Further analysis was then performed to localize the differences among the individual conditional means. As the aim of the experiments was to investigate the performance of additional depth information in comparison to the base condition, post-hoc comparisons were only conducted for all conditions versus the base condition to determine where significant differences existed. Seven paired sample t-tests were carried out on all possible combinations of other condition versus the base condition. For the t-tests the p value was adjusted for Bonferroni correction therefore $p < 0.007$ was required to achieve significance between the means of the conditions. Table 3 shows the results of the seven t-tests performed.

Table 3 shows that only the 3D movement condition, the co-ordinates display condition and the correct shadow condition versus the base condition reached the required $p < 0.007$ significance level. The mean difference for the 3D movement condition is significantly higher than the base condition; therefore the 3D movement condition mean is significantly less accurate than the base condition. However the Co-ordinate and Correct shadow conditions are significantly lower than the base condition and are therefore significantly more accurate than the base condition.

5 Discussion

The primary aim of this experiment was to investigate the beneficial affect of the presentation of additional depth information in a simple 3D AR target alignment task. Our initial hypothesis was that *any* additional depth information would aid a users performance over conditions without such depth information. Our results support this view; however they also suggest that certain depth cues are more helpful than others. In particular, co-ordinate information about the absolute position of the cursor and the addition of a correctly oriented shadow proved to significantly improve performance.

It is hypothesized that these results indicate the ability is for these experimental conditions to indicate the zero plane (when $Y=0$), thus affording the user useful information about the positional height of the cursor relative to its starting position. The benefit of this knowledge may be accentuated in our experiments as the $Y=0$ plane was situated on the tabletop, as were the bases of the target objects. Users' judgments must still therefore be made as an evaluation of the cursor position depending on the users understanding of the origins position. Enabling the user to

easily see when they have returned to this position, or the zero plane, increases a user's understanding of where the cursor is in relation to the perceived world.

Another significant finding highlights the benefit of control of individual dimensions of movement. All experimental conditions apart from the non-constrained 3D-movement condition allowed the user to switch on/off the individual x, y and z DoFs. Comparison of the base and non-constrained 3D-movement condition indicates the effect of the differing control techniques. As shown in Table 3, the base condition significantly outperforms the non-constrained 3D-movement condition. It can therefore be seen that having control over the individual x, y and z dimensions significantly increases the accuracy of performance during an alignment task. It is assumed that when in control of all dimensions; subjects are unable to distinguish possible errors made and therefore unable to reconcile them, therefore any increase in accuracy is due to the ability of users to limit errors to a single DoF.

6 Conclusions and Future Work

The results presented here show that many factors have an influence on the accuracy of object alignment, or selection, tasks in stereoscopic AR displays. Further work is required to determine the optimal combination of these influential factors. However, our experiments indicate that using additional depth cues such as simulated cursor shadows and text overlays of cursor positions significantly enhance subject accuracy in object selection. Additionally, allowing subjects to constrain the motion of the cursor also increases accuracy.

The results presented here have ramifications not only for the development of AR but any technology in which objects within a 3D world are required to be selected. In VE, interface techniques such as ray casting are designed to select entire objects, these techniques would have difficulty with the selection of more accurate or complex objects or parts of objects. Future desktops computers will inevitably utilize 3D space to cope with the ever-increasing multitasking workload of the user. Some 3D desktops have already been proposed [10].

Our experimental results also reinforce two well-known principles of human-computer interaction, namely feedback and control (see [11] for example). The non-constrained condition is worse than any other condition because users do not have adequate control whereas the three most accurate conditions provide control – either visually as in the shadow condition, or conceptually as the y co-ordinate shows zero in the co-ordinates display. As a general finding, we should conclude that developers of devices for use in AR worlds should provide better facilities for allowing users to take control and for providing suitable feedback as to position in the AR world.

It should also be noted that, of course, the use of desktop devices, such as the Spacemouse might not be suitable for manipulating cursors in wearable or mobile AR systems. Previously described systems of this kind have made use of hand-held joystick or glove type devices to perform basic object selection and menu driven commands. Although much work has been done in glove-based interfaces for VE interaction it is unclear whether such devices are appropriate for use in AR systems. Potentially new ways of implementing easy, fast and natural ways of object selection for walkable or wearable AR applications include eye movement analysis and even brain computer interface (BCI) techniques. Some work in eye-movement analysis for

non-cursor based selection of objects has already been described for VE's [12] – it is of interest to discover whether such techniques map easily to AR systems. The performance of BCI systems remains, at present, very limited – though it is intriguing to note that a good proportion of BCI work has been aimed at cursor manipulation (see [13] for instance). To-date such work has been limited to cursor control in one or two dimensions in purely computer generated environments– again it will be interesting to discover whether this can be extended to three dimensions and then, in turn, to AR environments.

Acknowledgments. This work has been funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant No. GR/R25309/01.

References

1. Azuma, R.; Bailiot, Y.; Behringer, R.; Feiner, S.; Julier, S. and MacIntyre, B. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, 25(6) (2001), 24-35.
2. Lawson, S.W., Pretlove, J.R.G., Wheeler, A.C., Parker, G.A. Augmented Reality as a Tool to Aid the Telerobotic Exploration and Characterization of Remote Environments. *Presence: Teleoperators and Virtual Environments*, 11(4) (2002), 352-367.
3. Brooks, Jr., F.P., What's Real About Virtual Reality? *IEEE Computer Graphics and Applications*, 19(6) (1999) 16-27.
4. Azuma, R.; A survey of augmented reality, *Presence: Teleoperators and Virtual Environments*, 6(4), (1997), 355-386,
5. Sands, J., Lawson, S.W. Towards a usable stereoscopic augmented reality interface for the manipulation of virtual cursors. To be presented at *ISMAR 2003, 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality*, Oct. 2003, Tokyo, Japan.
6. Pretlove, J.R., Lawson, S.W. Integrating Augmented Reality and Telepresence for Telerobotics in Hostile Environments, *Proc First IEEE International Workshop on Augmented Reality (IWAR'98)* A.K. Peters (1998) 101-107.
7. Milgram P., and Drascic, D. Perceptual Effects in Aligning Virtual and Real Objects in Augmented Reality Displays. *Proc. Human Factors and Ergonomics Society 41st Annual Meeting* (1997).
8. Piekarski, W., and Thomas, B.H. Tinmith-Metro: New Outdoor Techniques for Creating City Models with an Augmented Reality Wearable Computer. *Proc. 5th Int. Symposium on Wearable Computers*, (2001) 31-38.
9. Poupyrev, I., Weghorst, S., Billingham, M., Ichikawa, T. Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. *Computer Graphics Forum*, 17(3) (1998) 41-52.
10. Robertson, G.C.; van Dantzych, M.; Robbins, D.C.; Czerwinski, M.; Hinckley, K.; Ridsen, K.; Thiel, D.; and Gorokhovskiy, V. The Task Gallery: a 3D window manager, Proceedings of the SIGCHI conference on Human factors in computing systems, The Hague, The Netherlands. April, 2000, 494-501
11. Benyon, D. R., Turner, P. and Turner, S. *Designing Interactive Systems*. Addison-Wesley (2004).
12. Tanriverdi, V. and Jacob, R.J.K. Interacting with Eye Movements in Virtual Environments. *Proc. CHI 2000* ACM Press (2000) 265-272.
13. Black, M. J., Bienenstock, E., Donoghue, J. P., Serruya, M., Wu, W., Gao, Y. Connecting brains with machines: The neural control of 2D cursor movement. *Proc. 1st International IEEE/EMBS Conference on Neural Engineering*, (2003) 580-583.