STEREOSCOPIC VISION IN DESKTOP AUGMENTED REALITY
User Performance in the Presence of Conflicting Depth Cues

Gustavo Rovelo[1], Francisco Abad[2], M.-Carmen Juan[2] and Emilio Camahort[2]

[1]*Departamento de Sistemas Informáticos y Computación, Universitat Politècnica de València, Camino de Vera S/N 46022 Valencia, Spain*

[2]*Instituto Universitario de Automática e Informática Industrial, Universitat Politècnica de València, Camino de Vera S/N 46022 Valencia, Spain*

Keywords: Augmented Reality, Stereoscopic Vision, User Performance Assessment, Perceptual Issues.

Abstract: The use of stereoscopy as a depth cue in current computer graphics technology is increasingly popular. On the other hand, in a Desktop Augmented Reality (AR) environment, a web cam and a set of printed markers allow the user to interact with an augmented view of the environment in the computer display. So, it seems a logical step for Desktop Augmented Reality systems to take advantage of the available stereoscopic hardware to improve the realism of the augmented scenes. The main problem of Desktop AR applications is that they will be used in different and unprepared environments (wide variety of hardware, inconsistent and unpredictable lighting conditions, etc.) Therefore, it is crucial to understand how adverse conditions affect user experience. Furthermore, Desktop AR applications usually present conflicts in the depth cues presented to the users: viewpoint offset and incorrect occlusions between real and virtual objects.

We use the within subjects experimental approach to evaluate user performance in an Augmented Reality game. Our goal is to find if stereoscopic graphics help to reduce the impact of the other depth cues conflicts inherent to Desktop AR applications.

1 INTRODUCTION

Augmented Reality (AR) applications have become more common beyond the research laboratory. Users can now use AR systems at home while surfing the web (Sony Pictures Digital Inc., 2009), using their mobile phones (SPRX Mobile, 2010) or playing with a video console (Sony Computer Entertainment Europe, 2009).

The use of stereoscopy as a depth cue in computer graphics technology is increasingly popular too. However, stereoscopy is not the only depth cue humans use to perceive the third dimension. Motion parallax, perspective, accommodation, shadows, texture gradient and occlusion are other depth cues examples.

When implementing Desktop AR applications, it is common to introduce conflicts in different depth cues due to the limitation of the devices and the real-time constraints. For example, when occlusion between virtual and real objects is not correctly solved, virtual objects always appear on top of the real objects. The Figure 1 shows an example of this problem: the user's hand is occluded by the virtual object when the hand should be visible, since it is closer to the camera. These conflicting cues confuse the brain and reduce the realism of the application. Another problem of AR systems, that is magnified in Desktop AR environments is the viewing angle offset problem: real and virtual objects are shown in the display from the viewpoint of the web cam, that is quite different from the user's viewpoint (Kruijff et al., 2010). In professional AR systems, this problem is solved using see-through or video-see-through HMDs.

In this paper we study how users react to conflicting depth cues when completing a spatial task in a Desktop AR application. Particularly, we analyze if using stereoscopic graphics to render virtual objects increases the user's performance. Our testbed is a popular game for children, the Wire Loop Game. We compare user performance using both monoscopic and stereoscopic graphics.

We ran experiments with 32 volunteers of different ages. For each volunteer and each game, we collected five performance measurements. We then statistically analyzed the measurements to determine user performance. Additionally, each user answered

Rovelo G., Abad F., Juan M. and Camahort E..
STEREOSCOPIC VISION IN DESKTOP AUGMENTED REALITY - User Performance in the Presence of Conflicting Depth Cues.
DOI: 10.5220/0003852304600465
In *Proceedings of the International Conference on Computer Graphics Theory and Applications* (GRAPP-2012), pages 460-465
ISBN: 978-989-8565-02-0

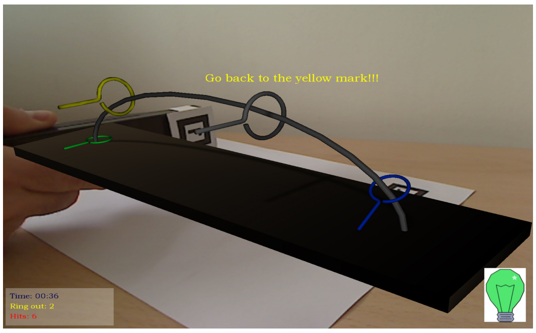

Figure 1: Virtual objects are always drawn in front of real objects. This figure shows the path for the easiest level of the game.

a questionnaire about non-quantifiable subjective issues about her game experience.

The outcome of the statistical analysis shows that using stereoscopic graphics does not benefit user performance in a Desktop AR application. We found that users performed better with monoscopic viewing. We also found that users playing with stereoscopy enabled first had worse results the second time they played (with monoscopic graphics) than users playing the game with a monoscopic view during the first round. This effect is even more noticeable when the task is more complex, as in the third (and last) level of our game.

The rest of the paper is organized as follows. First, in Section 2 we make a brief review of related work. Then, Section 3 describes the features of our AR game and system. In Section 4 we explain the goal of our research, the experimental design of our study and the user tests. Sections 5 and 6 describe and discuss our results. Finally, in Section 7 we present our conclusions and directions for future work.

2 RELATED WORK

Many authors have studied the problems that commonly arise in AR applications. There are surveys that discuss perception issues in AR systems like, for example, occlusion, stereoscopy, shadows, visuo-haptic collocation, user viewpoint tracking and visual angle offset between real and virtual objects (Azuma et al., 2001; Drascic and Milgram, 1996; Kruijff et al., 2010).

There are also more specific studies that analyze the impact of stereoscopic depth cues in AR systems (Lawson et al., 2002; Sands et al., 2004; Shidoji et al., 2010). Other works study the conflict of handling occlusion between real and virtual objects (Hou, 2003). Ruling *et al* propose a new algorithm to calculate the correct occlusion in unprepared scenarios where unmodeled real objects might hide virtual objects (Ruling et al., 2008).

Researchers have also studied how shadows can improve user's perception of depth and virtual objects (Sugano et al., 2003), or the minimal resolution in a light map to render perceptually correct artificial shadows (Nakano et al., 2008).

Other studies in user-centered evaluation focus on how improving the visualization hardware can alleviate AR perception issues (Cakmakci et al., 2004; Livingston et al., 2009; Liu et al., 2008).

All of the cited works try to understand and reduce perceptual issues in AR. Many times the solutions proposed use specialized tracking and rendering hardware. Other times only one or a few issues are considered instead of analyzing the whole picture. In our study we perform a user-centered evaluation to analyze the perceptual issues that arise in a Desktop AR application.

Our purpose is to evaluate whether stereoscopy helps or hinders the user's performance, taking into account the conflicting depth cues in this type of AR application. We quantify the user's performance by taking certain measurements described later. In the following section we describe the game we used to run the tests.

3 THE AR WIRE LOOP GAME

In the Wire Loop Game, players need psychomotor coordination and good depth perception skills. The goal is to move a wire loop along a wire path without the loop touching the path. If this happens, the game provides the user with some form of feedback. The AR Wire Loop Game we developed defines three different game levels. Each level has a different number of curves in the virtual wire path: one, three and five for the easy, intermediate and hard level, respectively. A Wii remote (Wiimote), the computer display and its speakers are used as the output devices to provide tactile, visual and aural feedback, respectively.

The virtual wire path and its wooden support are rendered on top of a real ARToolKit marker. We place the marker on a table in front of the system's camera and the stereoscopic display. The user can move and orient the virtual ring moving and orienting the Wiimote and its attached marker.

Two situations may arise while playing the game. First, the user may touch the wire with the ring. We call this situation a collision event. Second, a crossing event occurs when the user completely crosses the wire with the ring. In this case the game draws a yel-

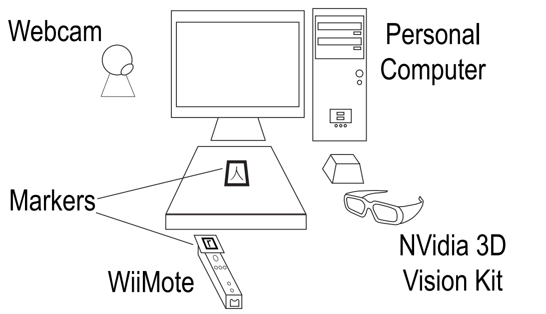

Figure 2: The components of our system: two printed markers, a PC with a NVIDIA Quadro 600 graphics card, a web cam, a Wiimote, a 3D display and the NVIDIA GeForce 3D Vision kit.

low virtual ring at that point (see Figure 1) and she will not be able to finish the level unless she takes the ring back at the point where the path was left in the first place. In both situations, visual, aural, and tactile stimuli are used to provide feedback to the user.

The Wiimote is connected to a PC using Bluetooth. We use the Wiiuse library to communicate our application with it (Michael Laforest, 2008). Stereoscopic graphics is achieved using an LG Flatron W23630 monitor and the NVIDIA GeForce 3D Vision Kit (active stereoscopic glasses and a NVIDIA Quadro 600 graphics card). Figure 2 illustrates our setup. We used osgART (Looser et al., 2006) to build our game. osgART is a toolkit that offers fast integration of virtual and real-world 3D objects.

Besides stereoscopy, our system implements other depth cues like shadows. All virtual objects, that is, the path and the ring, cast and receive shadows. We also use occlusion between virtual objects as a depth cue. Normally, part of the ring occludes the wire and vice versa (see Figure 3).

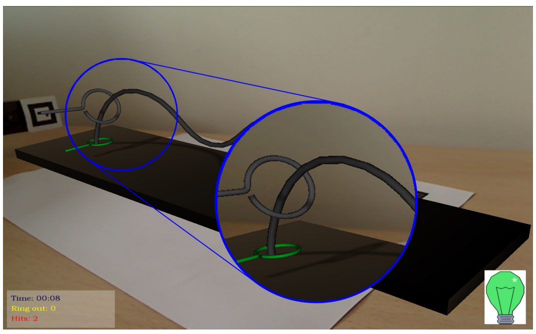

Figure 3: The ring and the path cast shadows over the virtual board and over each other. The occlusion is also solved between virtual objects.

4 EXPERIMENTAL DESIGN

We performed a study with 32 right handed volunteers aged 21 to 43 (27.5 on average). All of them volunteered their time. The design of our experiment defines two factors with two possible levels each. The first factor is using stereoscopic graphics (enabled or disabled). We use the Within Subjects experimental design, therefore all subjects played using stereoscopic and monoscopic graphics. The second factor of the experiment is the order used for each user (stereoscopic visualization first, then monoscopic; or the other way around). All users played the three levels of the game in the same order. Half of the users started with the monoscopic version, the other half with the stereoscopic version.

All subjects were sitted at about 50 cm from the monitor. The ARToolkit marker to track the wire path location was placed in the table in front of the monitor. The web cam was located on top of a tripod to the left of the user, to avoid interferences with her arms while playing the game. The game starts placing the loop at the farthest end of the path with respect to the user.

Before playing, each participant watched an introductory 3.5 minutes video with instructions on how to play the game. After that, a member of our team gave the user some final instructions. Then, the subject played the three levels (with stereoscopy enabled or disabled). After that, the subject played again the three levels, with the other configuration.

The first time a user plays, she spends one minute practicing with the first level of the game before running the experiments. The goal of each user is to finish each level in the shortest possible time and with the least amount of mistakes. Mistakes, as explained earlier, occur when the user touches the wire with the ring, and when the user crosses the wire with the ring. Users were allowed to take a one minute break between levels. Once all three levels were completed, we asked the users to fill out a questionnaire. We also gathered the user's opinion about the game by doing a brief interview with each one of them.

To perform our Within Subjects analysis, we recorded the following measurements: *Completion time* T (time required to complete a game level); *Number of collisions* N_c (number of times the user touches the wire with the ring); *Number of crossings* N_x (number of times the user crosses the wire with the ring); *Average collision time* T_c (average time the user takes to stop touching the wire after a collision) and *Average time outside the wire path* T_o (average time the user takes to bring the ring back to the wire path, after crossing it).

The times listed above do not include the times spent by our system generating the feedback stimuli. These times are negligible compared to the above measurements. For example, aural stimuli take 0.04 seconds on average to generate, while the Wiimote vibration requires 0.03 seconds on average to be activated. The worst-case scenario frame rate of our system occurs when stereoscopy and all three feedback stimuli are enabled. In that case the frame rate is 28 fps. The best-case frame rate is 36 fps.

5 STATISTICAL ANALYSIS

In this section we present the outcome of the statistical analysis of the results of our experiments.

To study the data of our tests, we did a Multifactor ANOVA analysis. We selected Fisher's Least Square Difference method to prove our hypothesis for all five measurements recorded and for both monoscopic and stereoscopic graphics. We applied the method at the 95% confidence level. We performed the analysis using StatGraphics® and assuming the following null and alternative hypotheses:

$$H_0 : \text{performance}_s = \text{performance}_m$$
$$H_a : \text{performance}_s \neq \text{performance}_m$$

Our H_0 states that using stereoscopic graphics will not improve user's performance in our Desktop AR testbed.

Table 1 presents the Multifactor ANOVA analysis outcomes for every one of the five measurements for the stereoscopy factor. We find a statistically significant difference only for T_o in Level 3; so we can not reject H_0.

Since we use the Within Subjects design, and every user plays each game level twice, we analyzed the relationship between both factors: experiment order and use of stereoscopy. We did not find any statistically significant relationship between both factors for any of the five performance measurements on any game level.

We detected the presence of outliers in the data. For this reason we performed a cleaning process to remove those samples. Despite this fact we can observe that the data dispersion in some cases is not small enough.

5.1 Questionnaires and Interviews

The tests we ran with our volunteers allowed us to gather subjective information too. We collected it using questionnaires and interviews. The questionnaires included yes/no questions, open questions and Likert scale questions (from one, strongly disagree, to seven, strongly agree).

The users' answers showed that more than 70% of them knew the real version of game. 56.5% of the users who knew the real version of the game liked our game more than the real version. Almost all users (84.3%) enjoyed playing it.

Even though users found the game easier without stereoscopy, 65.6% of them reported that stereoscopic graphics let them better perceive object depth. 25% reported that they perceived depth better without stereoscopic visualization and 9.4% reported no difference. These results were confirmed using a second question about the number of errors the users felt had incurred while playing: 28.3% of the user said they felt that stereoscopy had not helped them, 9.4% said that there was no difference and 62.3% that the stereoscopy helped them to had less mistakes.

We also asked the users about the game. They said that the stereoscopic version offered more information about the depths of the ring and the path. However, they complained that it was sometimes hard to accommodate their vision to focus on the objects, and that they were not always able to see 3D correctly. They also said to suffer ghosting (double images). Finally, camera and screen positions were mentioned to cause discomfort.

Conversely, the use of shadows as a depth cue was considered very useful. They were particularly useful at the beginning of the game, when moving the ring into the path for the first time.

6 DISCUSSION

In our experiments, each user played through the three levels of difficulty of the game twice, with and without stereoscopy. Each time we took five measurements: T, N_c, N_x, T_c and T_o. The first three measurements increase with the level of difficulty. This is independent of using monoscopic or stereoscopic graphics. This is due to the increased length of the wire and the increased number of curves in the wire.

Regarding the two reaction times, we observe a strange behavior for level 2. Users need more time to correct the ring position after touching the wire than in level 3. On the other hand, the use of monoscopic or stereoscopic graphics does not affect on average the times needed to correct the ring's position. As for user performance, we do not find statistical evidence to reject the null hypothesis for any measurement we took. This means that, even though on average monoscopic results are better than stereoscopic results, sta-

Table 1: Multifactor ANOVA analysis outcomes: stereoscopy factor.

Dependant Variable	Level 1	Level 2	Level 3
Completion time	$F(1,58) = 0.87, p = 0.3555$	$F(1,58) = 0.64, p = 0.4271$	$F(1,58) = 0.91, p = 0.3446$
Number of collisions	$F(1,58) = 0.24, p = 0.6233$	$F(1,62) = 0.87, p = 0.3561$	$F(1,62) = 0.33, p = 0.5684$
Number of Crossings	$F(1,58) = 0.03, p = 0.8742$	$F(1,58) = 0.20, p = 0.6573$	$F(1,60) = 0.02, p = 0.8970$
Avg. Collision time	$F(1,56) = 0.45, p = 0.5044$	$F(1,58) = 1.75, p = 0.1906$	$F(1,60) = 0.25, p = 0.6170$
Avg. Time outside the path	$F(1,58) = 0.54, p = 0.4660$	$F(1,60) = 2.22, p = 0.1418$	$F(1,60) = 4.06, p = 0.0485$

tistically using monoscopic or stereoscopic graphics does not affect user performance while playing the game.

Using stereoscopy in graphics applications improves depth perception, but our experiments show that under conflicting depth cues, stereoscopic graphics does not improve user performance in a position and orientation task like the one we evaluated. One reason is the fact that the user sees the virtual objects (and the real scene) on the display from the point of view of the camera, not from her own point of view (see (Azuma et al., 2001) for a detailed description of this problem). Another reason is that occlusion between real and virtual objects is not correctly handled. The user's hand and the Wiimote always appear behind the virtual objects. These contradictory depth cues confuse some users. They reported better depth perception, but the scene looked weird to them. Users had problems accommodating their vision to the 3D objects. Finally, some users reported eye strain after playing the game with stereoscopic glasses. All these perception issues are common to other AR applications, and have also been reported by (Kruijff et al., 2010).

Another reason why users performed better without stereoscopy is probably because they had enough information to determine the position and orientation of the virtual objects with just one image. The use of occlusion between virtual objects and the use of shadows as depth cues reduces the impact of conflicting depth cues. In unprepared AR environments, Sands *et al.* also showed that stereoscopy did not improve user performance in target selection tasks when there is enough information to find the position of a 3D cursor using a 2D visualization method (Sands et al., 2004).

Despite the fact that there are no statistically significant relationship between both factors, we performed a second statistical analysis. We compared user performance according to the type of visualization used first. We classified users into two groups (stereoscopic visualization first, monoscopic visualization first). If we take into account the average performance values, we find that in general both groups of users got better results in their second round. This improvement is minor for those subjects that started playing with monoscopic graphics, but is much greater for those who started playing with stereoscopic visualization. In general, all the measurements for the first and second rounds playing stereoscopic visualization first where worse than the first and second round playing monoscopic visualization first. Finally users playing with stereoscopy enabled first had worse results with monoscopic visualization condition in their second round than users playing the game with monoscopic condition as the initial round. Playing with stereoscopy enabled in the first round seems to penalize user performance in the second round, causing worse performance compared to users playing the game first using monoscopic graphics with no previous experience. This effect is more noticeable as the task complexity increase, from level one to level three.

Summarizing, our system in monoscopic mode provides in general enough cues for the user to correctly relate real and virtual objects. Consequently, users who started in this mode performed better on average. Also, the second time they play, using stereoscopy, they better interpreted the additional depth cues. Alternatively, when the user played with stereoscopy enabled first, she had to deal with the added complexity of the stereoscopy and the conflicting depth cues. Her performance was not as good until she played for the second time, in monoscopic mode, where there are less conflicting depth cues and it is easier to establish spatial relationships between real and virtual objects. These conclusions were pointed out by different users after completing their experiments.

We think that our results are application dependent and the lack of statistically significant difference that let us prove our hypothesis is caused for the small set and diversity of users that participated in our experiment. Therefore we plan to extend the experiment to include more users.

7 CONCLUSIONS AND FUTURE WORK

We have compared user performance when using stereoscopy in a Desktop AR system with conflicting depth cues (viewpoint offset and incorrect occlusions between real and virtual objects). We studied 32 volunteers that played three different levels of an AR version of the Wire Loop game. Each time we recorded five performance measurements.

We did not find statistically significant differences on user performance under both tested conditions. However, our results show that stereoscopy does not improve user performance in our Desktop AR application. Also, users playing first with stereoscopic graphics found the game easier after switching to monoscopic graphics. Users starting with monoscopic graphics had better results in all performance measurements.

Users completed a questionnaire after playing the AR game. They found the game easier to play without stereoscopy, although they said it helped them perceive the virtual objects better.

We plan to extend this study to include more users to get more data that support the findings we have described here. We also want to evaluate user performance when using other AR applications. Finally, we plan to study how different feedback channels may provide task related information to a user of an AR system.

ACKNOWLEDGEMENTS

This research was partially supported by the National Council of Science and Technology of México as part of the Special Program of Science and Technology. It was also funded by grant ALFI-3D, TIN2009-14103-C03-03 of the Spanish Ministry of Science and Innovation.
Finally, we want to thank Carlos De Jesús De Barros for his valuable criticism while preparing this paper.

REFERENCES

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. (2001). Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, 21:34–47.

Cakmakci, O., Ha, Y., and Rolland, J. P. (2004). A compact optical see-through head-worn display with occlusion support. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, ISMAR 2004, pages 16–25.

Drascic, D. and Milgram, P. (1996). Perceptual issues in augmented reality. In *SPIE Volume 2653: Stereoscopic Displays and Virtual Reality Systems III*, pages 123–134.

Hou, M. (2003). A model of real-virtual object interactions in stereoscopic augmented reality environments. In *Seventh International Conference on Information Visualization*, IV 2003, pages 512–517.

Kruijff, E., Swan, J. E., and Feiner, S. (2010). Perceptual issues in augmented reality revisited. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 3–12.

Lawson, S. W., Pretlove, J. R. G., Wheeler, A. C., and Parker, G. A. (2002). Augmented reality as a tool to aid the telerobotic exploration and characterization of remote environments. *Presence: Teleoperators and Virtual Environments*, 11:352–367.

Liu, S., Cheng, D., and Hua, H. (2008). An optical see-through head mounted display with addressable focal planes. In *7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR 2008, pages 33–42.

Livingston, M. A., Barrow, J. H., and Sibley, C. M. (2009). Quantification of contrast sensitivity and color perception using head-worn augmented reality displays. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pages 115 –122.

Looser, J., Grasset, R., Seichter, H., and Billinghurst, M. (2006). OSGART - A Pragmatic Approach to MR. In *Industrial Workshop*, ISMAR '06.

Michael Laforest (2008). Wiiuse - The Wiimote C Library. http://sourceforge.net/projects/wiiuse. Last checked: October 7th, 2011.

Nakano, G., Kitahara, I., and Ohta, Y. (2008). Generating perceptually-correct shadows for mixed reality. In *7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR 2008, pages 173 –174.

Ruling, Z., Hanxu, S., Qingxuan, J., and Fusheng, Y. (2008). Research on fast and accurate occlusion detection technology of augmented reality system. In *Industrial Informatics, 2008. INDIN 2008. 6th IEEE International Conference on*, pages 111–116.

Sands, J., Lawson, S. W., and Benyon, D. (2004). Do we need stereoscopic displays for 3D augmented reality target selection tasks? In *Eighth International Conference on Information Visualisation, 2004.*, pages 633–638.

Shidoji, K., Funakoshi, M., and Ogawa, M. (2010). Perception of absolute and relative distances in stereoscopic image. *Displays*, 7524:752418–752418–10.

Sony Computer Entertainment Europe (2009). Invizimals. http://www.invizimals.com/invizimals.php. Last checked: October 7th, 2011.

Sony Pictures Digital Inc. (2009). District 9 Movie. Multi-National United Training Simulation. http://www.multinationalunited.com/training/. Last checked: October 7th, 2011.

SPRX Mobile (2010). Layar. http://www.layar.com/. Last checked: October 7th, 2011.

Sugano, N., Kato, H., and Tachibana, K. (2003). The effects of shadow representation of virtual objects in augmented reality. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003.*, pages 76–83.