A framework for comparing task performance in real and virtual scenes

Sarah Howlett, Richard Lee, Carol O'Sullivan*
Image Synthesis Group, Trinity College Dublin

Abstract

In this paper, we describe a framework for comparing task performance in real and virtual environments. Realistic graphics, rear projection, haptics and rapid prototyping are used to match the virtual scene to the real scene. We describe some preliminary placement tasks which were evaluated using eye-tracking and discuss our future plans for this framework.

CR Categories: I.3.6 [Computer Graphics]: Methodology and Techniques–*Interaction Techniques*; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

Keywords: visual perception, attention, task performance, real and virtual environments

1 Introduction

Often it is necessary for computer graphics to operate under realtime constraints as well as maintaining realistic and dynamic scenes. Therefore it is useful to know what factors influence perception and can be allocated more resources at the expense of other aspects. In previous research we attempted to determine salient features of objects using eye-tracking. That research demonstrated that, if the salient features of natural objects are preserved during simplification, the visual fidelity of these models can be improved [Howlett et al. 2004]. However, we found that visually prominent features for man-made artifacts were not so easy to predict. A reason for this may be that man-made artifacts are generally related to a task and that prominent features may be defined by this task.

It is commonly known that the task performed on objects plays an important role in determining where attention is focused [Hayhoe et al. 2003], which leads us to the next step in our research. There are many factors that define a task, such as the environment, the nature of the task and the objects involved in the task. There has been much research in the field of psychology, examining eye-movements during a 3D task. However, there are further differences in performance between real and virtual environments.

The aim of our current work is to investigate these issues further by comparing tasks in the real and virtual world. To this end, we are building a framework using realistic graphics, rear projection, haptics, rapid prototyping and eye-tracking for evaluation. In this framework we attempt to replicate as accurately as possible a real world scene which we have created and the interaction with this world using haptics. Future plans for our framework include; investigating how attention is captured for a variety of tasks, determining the differences in performance and strategies between real and virtual situations, and finding the limitations of carrying out similar tasks in a virtual environment. Our aim is to determine how the user experience can be improved, perhaps through previewing or by displaying salient object features [Howlett et al. 2004] or task related objects [Cater et al. 2003] in greater detail.

This is work in progress, and the computer framework is not complete. We have carried out some preliminary studies on the real world scene and its virtual counterpart. The remainder of this paper describes the implementation of our framework and our efforts so far to match the two setups as accurately as possible, some preliminary tests and observations, followed by a discussion of our future plans for this framework.

2 Background

There is much research from the field of psychology suggesting that visual activity is largely controlled by task. Hayhoe et al. [2003] recorded participants' eye-movements while making a sandwich, and noted that almost all of the fixations focused on the task. Supporting this is work studying the eye-hand coordination in object manipulation [Johansson et al. 2001]. It was demonstrated that the salience of gaze targets arise from the functional sensorimotor requirements of the task, and therefore participants directed gaze almost exclusively towards objects involved in the task. Pelz et al. [2001] monitored participants' eye-movements during the familiar complex task of hand washing, which revealed a novel perceptual strategy involving task-dependent lookahead fixation. More recently, Ling et al. [2004] showed that color and size interact in a real 3D object similarity task.

From the field of computer graphics, recent work by Cater et al. [2003] also supports the suggestion that visual attention is largely controlled by the task. They show experimentally that it is possible to render scene objects not related to the task at a lower resolution without the viewer noticing any reduction in quality. Sundstedt et al. [2004] investigate to what level viewers fail to notice degradations in image quality, between task related areas and non-task related areas. Watson et al. [2003] investigated the relationship of delay to user performance in 3D object placement. They showed that delay has a greater impact on performance when difficulty is high. Furthermore, they demonstrated that, with previewing, more delay can be tolerated even when the task is difficult.

There have also been previous studies comparing performance in real and virtual setups. Thompson et al. [2004] have shown that participants are significantly less accurate at judging distances in visually immersive environments than in the real world. Moreover, Mohler et al. [2004] carried out studies using treadmill-based virtual environments to simulate the perceptual-motor effects of actually walking around in the real world. In task related research, Lok et al. [2003] pointed out that having interactions with real objects within a virtual world can increase the effectiveness of the environment.

In our research we wish to expand upon some of these approaches, by examining task performance in a truly interactive, multisensory environment. As described in the literature, we want to examine the eye-movements of participants while they carry out various 3D

---

*e-mail: Sarah.Howlett, Richard.Lee, Carol.OSullivan@cs.tcd.ie

tasks in a real world situation, but we also extend this idea by comparing the eye-movement results to those found for a similar task carried out in a virtual environment.

3 Framework

3.1 Real environment

The real scene consists of a five-sided box, painted matte white, of dimensions 90x90x90 cm, as described by Morvan and McNamara [2003]. The box is divided equally into three regions using two horizontal shelves and is placed on a table 73 cm above the ground (see Figure 1 of the color plate). Participants sit on an adjustable chair in front of the box so that their eye level and field of view may be controlled. The environment is lit by a single 150 W bulb mounted above a square 7x11 cm opening in the top of the box. The two shelves are partitioned at the center so that each box region has a different illumination level. Directing the light in this way makes it easier to realistically model the lighting conditions on the computer (harder shadows, less indirect light, etc...).

Participants *may* interact with the scene by manipulating a selection of physical models, sourced from freely available data such as the Stanford bunny and Utah teapot. The models were fabricated using a Dimension 3D printer and painted in varying shades of matte grey to increase contrast with the background environment and provide a range of luminances. Plastic hooks were affixed to the top of these models to allow them to be picked up and repositioned with a 54 cm long plastic hand-held rod. We chose this means of interaction as it resembles the sensation experienced when lifting virtual objects with the haptic input device we are using in the virtual environment.

3.2 Virtual environment

The experience of interacting with the real environment is created on computer by back-projecting an OpenGL application onto a Filmscreen 150 canvas using a high quality DLP projector (see Figure 2 of the color plate). In an attempt to match accommodative and vergence distances between the real and virtual environments we place this screen the same distance from the participant as the physical box in the real world experiments, and adjust the size of the projected box so that it matches the dimensions of the front face of the physical box. Models of the box and rod were reproduced in Autodesk 3ds Max and the rest of the objects were rescaled to match their 3D-printed, real world counterparts, and appended with hook models.

The scene is viewed from a fixed camera pose, chosen to match the participant's required viewpoint when performing the real world experiments. Since the box and the surrounding environment remain static for the duration of the experiment, we decided to use a background image in place of an OpenGL rendering of the box for increased realism. Note that we still render the box geometry using OpenGL in order to update the depth buffer so that the object models are correctly sorted into the background plate.

To capture the image, a digital camera was matched to the virtual camera pose and a high dynamic range photograph of the box was constructed in HDRShop. This was then converted into a tone-mapped low dynamic range image to account for the detail visible to the human eye under the extreme range of brightnesses present in the real environment.

The dynamic scene elements are lit using a single, downward-pointing, OpenGL directional light, appropriately attenuated to account for the differing light levels in each box region. Simple, yet effective, shadows are rendered by plane projecting model geometry onto the shelves and fading their intensity based on height.

Participants interact with the scene using a Phantom Premium 6DOF device to manoeuvre the rod in 3D (see Figure 3 of the color plate). Currently, an object may be “picked up” by pressing and holding the Phantom switch when the rod is within range of the objects hook. Once selected, participants feel inertia, weight and collision forces on the Phantom. Due to the difficulty of computing stable haptic forces between groups of arbitrary meshes in contact, we chose to simplify the collision detection and force feedback problems by treating all objects as axis-aligned boxes, including the geometry of the five-sided box.

Haptic forces are computed by treating the selected object as a dynamic rigid body to which weight, damping and penalty-based collision forces are applied. Under the assumption of constant density, we set the mass of each object proportional to its computed volume. Collision detection is performed by analytically computing intersection times between static boxes and swept boxes. Even though the objects are always drawn in a collision-free position, internally, the boxes are allowed to intersect each other in order to provide penetration depth information from which stable haptic forces can be computed. The Phantom feedback force that results from movement of the dynamic body is calculated by coupling the body's center of mass to the Phantom position with a simulated spring and setting the feedback force proportional to the spring tension force. This model accurately simulates the sensation of lifting and dragging objects with the rod in the real world.

4 Preliminary experiments

4.1 Eye-tracker

During the experiments, the EyeLink II eye-tracker with scene camera was used to obtain eye-movement data. This allowed eye movements at different viewing depths to be tracked with high accuracy in the same recording, meaning participants could move their heads freely. Prior to each experiment, the eye-tracker and scene camera setup was performed including scene camera alignment, display area detection, calibration and depth correction.

4.2 Procedure

Eight people participated in the experiments, 4 for each scenario, real and virtual, while wearing the eye-tracker. Four of the participants were seated in front of the real environment on an adjustable chair and given a trial run of moving objects around using the rod. Similarly, in the virtual case participants were seated in front of the display, and instructed to practice moving the objects around using the Phantom. Each participant carried out two placement tasks. In one task, participants had to organize the models on the shelves according to luminance. They were given the instruction to place the dark shaded objects on the top shelf, light shaded objects on the bottom shelf and the medium shaded objects on the middle shelf. In the second task they had to arrange the objects depending on their type. They were told to place all natural objects on the left side of the middle shelf and man-made artifacts on the right side. Each version of a task had the same number of models arranged in a random order.

TASK	SETUP	TASK DUR	FIX DUR	SAC AMP
Luminance	real	77secs	264msec	6.1deg
Luminance	virtual	136secs	353msec	6.5deg
Object type	real	53secs	284msec	4.8deg
Object type	virtual	79secs	403msec	4.5deg

Table 1: Average results over all participants for the task duration, saccade amplitude and fixation durations during the trials.

The data was analyzed using the EyeLinkDV software which gave information regarding fixations and saccades. Furthermore, the video footage from the the scene camera, which included an overlay of each participant's gaze position, was examined.

4.3 Preliminary Results and Discussion

The tasks took longer to perform in the virtual environment, which is not surprising. The number of saccades and fixations could not be compared, as the trials took different lengths of time. However we found interesting results for the average fixation duration. In both tasks, the average fixation duration was longer in the virtual setup. In the real and virtual environments, the average saccade amplitude was greater for the luminance task than the object tasks for all participants. This demonstrates that the nature of the task has a similar influence in the real and virtual setup (see Table 1).

Some patterns were identified in the eye-movement data of the majority of participants during the examination of the scene camera data. In the real world setup, we also found evidence of look ahead fixations as in [Pelz et al. 2001]. This occurs when objects of future interactions were foveated before they were needed, i.e. when the next object to be picked up was fixated during the placement of the current object. It appears that, prior to placement, participants look ahead for free space to place the current object down. Surprisingly, it appears that the majority of participants fixate on the objects on either side of the free space and not on the free space itself. Furthermore, if a possible collision object can be picked up next, participants generally go for it with little or no regard for other pick-up options. They do not examine if there is free space for an object until after it is picked up. There is relatively little fixation on the rod, as the objects seem to get the majority of attention. When placing an object, participants perform advance planning, by fixating the next object they will pick up. There are some differences between participants, as some are more efficient and some people tend to look around. Most people seemed to adopt a strategy of picking up objects sorted by distance from the previously placed one.

The biggest difference between eye-movements in the real and the virtual setup is that, in the latter, fixations are far more concentrated on the object that the participant is currently moving, compared to the real world where they generally look around when the objects are on the rod. As a result of this, their eye-movements follow the objects that they are moving in the virtual world. For this scenario the tendency is not to plan ahead, but to deal with one object at a time for the task in hand. This is supported by the fact that the average fixation duration was greater in the virtual world.

Since the fixation patterns show that there is a difference between the real and virtual setup, we need to come up with ways of making it easier for the user in the virtual environment. It is important to note that results reported here are only from preliminary studies, to provide some quick insights. As the study uses a between-subject design and the number of participants involved was too small we did not perform any statistical significance tests on the results. In the future we will certainly repeat the study on a bigger group and perform a proper statistical evaluation.

5 Future Work

Our framework can be improved in a number of ways. Firstly, we experienced a major problem using the eye-tracker's scene camera with a single-chip DLP projector. These projectors beam white light through a spinning color wheel which filters it into red, green and blue components sequentially in time. Strobing effects result due to the insufficient sampling rate of the scene camera. We temporarily overcame this problem by instead using an overhead LCD projector. Note however that rear projection is greatly preferred to front projection as it allows participants to be seated directly in front of the virtual scene. Therefore, a better solution to this problem is to instead use a three-chip DLP projector which should prevent color cycling since red, green and blue color components are displayed simultaneously.

The most important discrepancy between our current real and virtual environments may be the lack of stereo vision in the virtual environment. We would like to address this by augmenting our framework with a rear projection stereoscopic display which will allow us to measure the effect of stereopsis on task performance in virtual environments. We have recently installed a Vicon optical motion capture system and will use this to examine motion parallax effects by tracking the head movement of participants, thereby allowing a dynamic viewpoint in our software. The influence of sound on task performance is another area of interest.

It is difficult to find a means of interacting with real world objects which can be faithfully reproduced in the virtual world. Options include the use of data gloves and motion capture. Hand [1997] examines some interaction techniques which have been developed for object manipulation, navigation and application control in 3D virtual environments. However, we chose the hand-held rod interface because it closely mimics the feel of the Phantom device when physically accurate feedback forces are applied to it, and we are very interested in examining the influence of haptics on task execution, including the use of non-physical forces to help "guide" the user during tasks. However, further tests to assess the faithfulness of the Phantom device in replicating the sensation of using the rod are needed.

Additionally, in the current set of experiments, the participants were seated directly in front of the scene, with the haptic device located to their right. This seemed the natural place to put it given their position and that they were all right handed. However, perhaps it would have felt more like the real world experience if it had been placed directly in front of the user. This will be considered in future experiments.

From an evaluation perspective, more experiments need to be carried out on the framework. So far, we have only examined eye-movements for a small number of participants during two placement tasks. Future experiments on the framework will involve more participants carrying out a wider variety of tasks e.g., passive, counting and memory, in addition to placement tasks.

The interesting observation about average fixation length merits further study. Perhaps in the future this could be used as a measure to compare how similar the real and virtual setups are. As more attention to the current object is needed to perform the same task in the virtual environment, this demonstrates that we need to come up with ways to facilitate the user. Possibilities would be to determine if enhancing the salient features of objects [Howlett et al. 2004], or like Cater et al. [2003], enhancing the task related objects could make it easier for the participant. Examining practices such as priming through previewing various aspects of objects would also be interesting. For example, Olds and Fockler [2004] compare the previewing of color and orientation, and show that conjunction previewing is most effective. Furthermore, Watson et al. [2003] demonstrated that, with previewing, more delay can be tolerated even when the task is difficult. Finally, the user's attention could be guided during a task. Halper et al. [2005] show that non-photorealistic rendering can play a subtle, yet effective, role in guiding judgement and subsequent interactions. We intend to examine whether additional cues can be used to augment the virtual environment, resulting in improved task performance.

References

CATER, K., CHALMERS, A., AND WARD, G. 2003. Detail to attention: Exploiting visual tasks for selective rendering. In *Proceedings of the 2003 EUROGRAPHICS Symposium on Rendering*, EUROGRAPHICS, P. Christensen and D. Cohen-Or, Eds., 270–280.

HALPER, N., MELLIN, M., DUKE, D., AND STROTHOTTE, T. 2005. Implicational rendering: Drawing on latent human knowledge. In *ACM Transactions on Applied Perception (to appear)*.

HAND, C. 1997. A survey of 3-d interaction techniques. In *Computer Graphics Forum*, vol. 16(5), 269–281.

HAYHOE, M. M., SHRIVASTAVA, A., MRUCZEK, R., AND PELZ, J. 2003. Visual memory and motor planning in a natural task. In *Journal of Vision*, vol. 3, 49–63.

HOWLETT, S., HAMILL, J., AND O'SULLIVAN, C. 2004. An experimental approach to predicting saliency for simplified polygonal models. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, ACM Press, 57–64.

JOHANSSON, R. S., WESTLING, G., BCKSTRM, A., AND FLANAGAN, J. R. 2001. Eye-hand coordination in object manipulation. In *The Journal of Neuroscience*, vol. 21(17), 6917–6932.

LING, Y., AND HURLBERT, A. 2004. Color and size interactions in a real 3d object similarity task. In *The Journal of Vision*, vol. 4, 721–734.

LOK, B., NAIK, S., WHITTON, M., AND BROOKS, F. P. 2003. Effects of handling real objects and avatar fidelity on cognitive task performance in virtual environments. In *Proceedings of the IEEE Virtual Reality*.

MOHLER, B. J., THOMPSON, W. B., CREEM-REGEHR, S., PICK, H. L., WARREN, W., RIESER, J. J., AND WILLEMSEN, P. 2004. Visual motion influences locomotion in a treadmill virtual environment. In *APGV '04: Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, 19–22.

MORVAN, Y., AND MCNAMARA, A. 2003. Assessing the visual perception impact of indirect lighting under different levels of illumination. a psychophysical experiment. In *Proceedings of Eurographics Ireland*.

OLDS, E. S., AND FOCKLER, K. A. 2004. Does previewing one stimulus feature help conjunction search? In *Perception*, vol. 33, 195–216.

PELZ, J. B., CANOSA, R., BABCOCK, J., AND BARBER, J. 2001. Visual perception in familiar, complex tasks. In *Proceedings of the ICIP*, vol. 2, 12–15.

SUNDSTEDT, V., CHALMERS, A., CATER, K., AND DEBATTISTA, K. 2004. Top-down visual attention for efficient rendering of task related scenes. In *Vision, Modelling and Visualization*, 209–216.

THOMPSON, W. B., WILLEMSEN, P., GOOCH, A. A., CREEM-REGEHR, S. H., LOOMIS, J. M., AND BEALL, A. C. 2004. Does the quality of the computer graphics matter when judging distances in visually immersive environments? In *Presence: Teleoperators and Virtual Environments*, 560–571.

WATSON, B. A., WALKER, N., AND WOYTIUK, P. 2003. Maintaining usability during 3d placement despite delay. In *VR '03: Proceedings of the IEEE Virtual Reality 2003*, 133–140.