Literature Review

JOSHUA RAMSBOTTOM, University of Cape Town

In recent years there has been an increase in graphical processing power of hardware and this has lead to an increased demand for 3D virtual models for animated films. The process of previsualization is a crucial part of the pre-production phase of a film. This process requires efficient 3D tasks to be performed within scenes of the film. Previous research has shown that 3D user interfaces are more effective for performing 3D tasks than the traditional approach. The recent surge in popularity of commercially available head-mounted virtual reality displays and their relatively low cost make HMD-based VR a good candidate for a 3D previsualization interface. 3D interfaces create a new set of problems for user interaction as traditional methods do not perform well in 3D space. A large body of research exists regarding new solutions to 3D user interaction. Based on recent research a hybrid approach using both 3D gestures and 2D touch input is promising for a HMD-based VR previsualization system.

1. INTRODUCTION

In the animated film industry previsualization (Previs) is a process that happens during the pre-production phase of a film. This technique is used to visualize and plan the camera work over the course of the film before the production phase. Many programs originally designed for 3D graphics are now used for previs. These programs are often require training to use. Previs also requires 3D tasks be performed such as manipulating objects in 3D space, but interaction with these traditional programs is limited by 2D input devices such as a mouse, and a 2D interaction window. This means that the previs process can often be tedious for what is meant to be a low-fidelity planning process. Nitsche [2008] argues that the interactive aspects of games combined with the way movies simply tell stories make game engines suitable tools for previs. He also argues that current previs software does not allow users to play a character in the film. This makes the combination of a game engine with virtual reality (VR) interaction a suitable candidate for solving this problem.

2. VIRTUAL REALITY

While there is no previs software that makes use of VR for interaction; there are existing implementations of 3D modeling software that make use of VR. These can be used as a proof of concept as the tasks performed in modeling software also involve manipulating objects in a 3D space, which is similar to previs. Butterworth et al. [1992] developed a 3D modeling tool that makes use of a head-mounted display (HMD). They support their decision to use a HMD by arguing that the HMD allows the user to be placed in the scene which makes it easier to understand the spacial relationships between models. The head tracking that comes with the HMD also allows the user to make fine adjustments to his/her viewpoint in a natural way.

Yang et al. [2014] found that users completed the task of moving an object in 3D space faster when using a system with stereoscopic vision and hand gestures as opposed to a mouse. Hughes et al. [2013] built CaveCAD, which makes use of 3D interaction in a virtual CAVE environment for building architectural models. They found that users required previous knowledge of modelling techinques to make use of the system effectively. Experienced 3D modellers were able to recreate an entire building in under 15 minutes. Ponto et al. [2013] created SculptUp, another CAVE-based 3D interaction modelling system. Users were able to create complex scenes in less than 10 minutes. Wang and Lindeman [2014] developed a level editing system that makes use of a non-oclusive HMD along with 2D tablet-based input. The tasks required in level editing are similar to those required in previs.

While not all of these implementations make use of HMD-based VR, there is common evidence that performing 3D tasks within a 3D virtual environment using 3D input devices leads to better task performance and efficiency [Butterworth et al. 1992; Yang et al. 2012; Hughes et al. 2013; Ponto et al. 2013; Wang and Lindeman 2014].

With the recent surge in consumer-level VR HMDs such as the *Oculus Rift* this technology is becoming more widely available at a more affordable cost. This, along with the increasing power of consumer desktop hardware and free 3D modeling and animation software like *Blender* and *Maya*, as well as game engines which support film production (or *machinima*) are making it possible for more people to create their own animated films [Nitsche 2008].

There are also problems regarding interaction with 3D virtual environments (VE). Most of these problems involve menu navigation and numeric data entry [Mine et al. 2014; Wang and Lindeman 2014]. This is caused by the lack of 2D pointing precision - which is required to use menus effectively - of 3D input devices.

3. 3D INTERACTION TECHNIQUES

Interaction with 3D virtual environments creates a whole set of different interaction metaphors. These metaphors are used to describe how a virtual environment will respond to user interaction by likening it to a real-world physical response. Jankowski [2011] provides a taxonomy of general tasks that are performed in 3D virtual environments from which these metaphors can be derived. While these tasks were classified with the 3D Web in mind, they still apply to VR.

Navigation - this involves the user changing his/her viewpoint within the VE to see different parts of it. Navigation is subcategorized into *General movement* - which is exploratory movement without any particular goal; *Targeted movement* - which is movement with regard to some specific point of interest, often to examine an object; *Specified coordinate movement* - which is movement to a single position and orientation; *Specified trajectory movement* - which is movement along a specific set of positions and orientations. Chittaro and Burigat [2004] investigated the effect of using 3D arrows as an aid to VE navigation. They found that while 3D arrows are only at least as effective as 2D aids, they excel in aiding 3D navigation with ”flying” movement. Ware and Osborne [1990] developed three ways for a user to move his/her viewpoint within a VE using physical objects . These three methods stemmed from three navigational metaphors that are defined later in this section. Tollmar et al. [2004] created a computer vision based navigation system where users interacted either via pointing gestures or direct full body movements. They found that these methods of interaction were more natural and thus more immersive. When tested against typical keyboard interaction however, the time taken to perform navigational tasks in a VE was higher. This was attributed to variances in each user's movements and no ability to change between fine and coarse movements.

Wayfinding - this describes the methods users use to create a mental model of a VE and is mostly used to supplement navigational tasks.

Selection - this is the process that occurs when a user wants to designate which object in a VE will be selected for navigation or manipulation. In 3D VEs a common metaphor for selection is *pointing*. A small body of research exists on direct 3D point selection techniques [Stuerzlinger and Teather 2014]. The majority of these techniques fall under two paradigms: ray-based techniques and virtual hand techniques [Stuerzlinger and Teather 2014; Teather and Stuerzlinger 2014]. The main difference between these paradigms is that virtual hand techniques check for intersection between the hand/cursor and thus require depth information. Dang [2007] provides a comprehensive classification of 3D pointing techniques, and adds a third major paradigm: spotlight. Spotlight-based techniques are similar to ray-casting techniques, but instead a conical projection is used to select objects. Lubos et al. [2014] developed a virtual hand technique that uses a hand-mounted 6 DoF infrared tracking target to capture hand movements, arguing that virtual hand interaction is more natural. In their study participants had to perform 3D selection tasks while wearing a HMD and evaluated using Fitts' Law. It was found that a significant amount of errors occured along the viewing direction while changes in the movement direction did not cause a significant amount of errors. They also propose two guidelines for HMD environments. The first guideline stipulates that 3D selection tasks which require fine movement should be restricted to objects that are close to the eyes. The second guideline stipulates that for selecting points in a 3D space, an ellipsoid cursor should be used in order to enlarge the selection space and reduce errors. Grossman and Balakrishnan [2006] found that a ray-based cursor performed much better than a 3D point cursor (virtual hand) for directly selecting an object in a sparse environment on a 3D volumetric display. They also developed an augmentation to the ray cursor, namely a *depth ray*. This adds a small sphere which can be moved dynamically along the ray to represent depth. This was found to be effective in counteracting the difficulty of selecting objects in a cluttered scene where some objects occlude others. Bowman et al. [1999] also found that a ray-based selection method significantly outperformed a 3D point cursor based method for 3D selection tasks. Teather and Stuerzlinger [2014] found that highlighting targets both increased movement speed and decreased the error rate of 3D point selection tasks.

Manipulation - this refers to methods used to specify a specific object's position, orientation, and scale. Manipulation is subcategorized into *Using a Manipulator* - virtual handles are attached to the object which are used to manipulate various aspects of the object, these handles are displayed together with the object; *Automatic Viewing Control* - the position of the virtual camera is used to augment manipulation; *Constrained Manipulation* - manipulation is simplified by using knowledge of physical aspects of the real world and objects to constrain it. In a constrained virtual assembly environment, Zhang et al. [2005] found that the addition of both 3D visual and auditory feedback improved assembly task performance. Gallo et al. [2008] developed a 3D interface for manipulating 3D medical scans using a Wiimote and voice commands to make common medical tasks more efficient. The Wiimote was chosen as it facilitates both 3D gestures and discrete input via buttons on the device. The 3D gestures were used to transform the object within different modes. Users were able to change between modes using the buttons. Laundry et al. [2010] tested the performance of four interaction methods for 3D object manipulation tasks. They found that an infrared pen with a click button (similar to a laser pointer, except the button is used to indicate a ”click”) performed the best, this is similar to the ray-based selection techniques described earlier. Recent research has shown a trend in hybrid interaction techniques. Wang and Lindeman [2014] developed a virtual level editing system using a non-occlusive HMD combined with a tablet. The HMD provided a first person view of the scene while the user was still able to view the tablet, which provided an overview of the scene. They found that this approach leveraged the strength of both devices and in turn, could lead to good task performance and user experience. It was also found that users preferred it when tasks were synchronized across the devices. Mine et al. [2014] created hybrid controllers that consisted of a smartphone attached to a 3D printed shell housing a microcontroller and three physical buttons. The smartphone was used for 2D touch input and the microcontroller facilitated 3D gesture input by moving the hybrid controller in physical space. The physical buttons were reserved for fundamental actions required at all stages of use. It must also be noted that both these studies of hybrid interfaces also require some amount of training before they can be used effectively.

System control - this refers to interactions between the user and the system which are not represented in the VE, such as changing the mode of interaction. Gebhardt et al. [2014] developed a system that uses a smartphone for system control menu interaction. While experts approved of the system, user acceptance was lower than expected.

Ware and Osborne [1990] define three metaphors for 3D user interaction, namely Eyeball-in-Hand, Scene-in-Hand, and Flying Vehicle Control. In the Eyeball-in-Hand metaphor the user's hand movements translate directly to viewpoint movements. It was found that users engaging with this metaphor had to consciously calculate movements. It was also noted that this can disorientate users, and that an overall plan view of the scene should also be supplied. In the Scene-in-Hand metaphor, the user's hand movements translate the entire scene. It was found that this metaphor can be useful for manipulating specific objects and changing the viewpoint in a scene. In the Flying Vehicle Control metaphor, the user's interaction controls a virtual vehicle to make it move through the scene.

4. DISCUSSION

While the previous work mentioned in this review is not exhaustive, there is a common theme among the findings. Many novel solutions have been developed and while they are all different, together they show the advantages and disadvantes of 3D versus 2D interaction methods. Selection and manipulation are very closely linked and studies that test 3D object manipulation tasks also involve selection. These two interaction metaphors are important for previsualization as they are required to perform the fundamental task of manipulating objects in a 3D scene. The work discussed in this review has shown that 3D gesturing - either by using a physical device or capturing hand movements - is more effective for these 3D selection and manipulation tasks. It has also been found that it is difficult to navigate menus and enter data using 3D input. Previous work as also shown that 2D touch interaction (using a smartphone or tablet) performs well in menu navigation tasks. This is a logical conclusion as menus were originally designed for 2D displays.

The exception to this is 3D navigation. While 3D visual aids in the VE improve navigational task performance, 3D input methods for navigation are not always faster than traditional 2D input using a keyboard and mouse.

It must also be noted that due to the novelty of the solutions developed in previous research, usability studies may not be very extensive. A smaller sample size and covenience sampling are used to establish the initial potential of a solution.

5. CONCLUSIONS

In this review I have summarized and discussed previous research on VR for 3D applications, as well as 3D user interaction metaphors that will guide the interaction process for a 3D previsualization system. The potential for a HMD-based VR solution is interesting as it provides the user with an immserive view of a scene, and even allows one to view the scene from the perspective of a character within it. This could lead to a more intuitive and effective means of previsualization. There are also challenges that come with this approach, particularly regarding user interaction. Given recent research a hybrid approach is promising. Providing the user with the ability to interact using 3D gestures and 2D touch with a combination of devices will help overcome the difficulties with 3D user interaction. As for the task of 3D navigation within a scene the preferred interaction method is still unclear.

REFERENCES

Doug A. Bowman, Donald B. Johnson, and Larry F. Hodges. 1999. Testbed Evaluation of Virtual Environment Interaction Techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '99)*. ACM, New York, NY, USA, 26–33.

Jeff Butterworth, Andrew Davidson, Stephen Hench, and Marc. T. Olano. 1992. 3DM: A Three Dimensional Modeler Using a Head-mounted Display. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics (I3D '92)*. ACM, New York, NY, USA, 135–138.

Luca Chittaro and Stefano Burigat. 2004. 3D Location-pointing As a Navigation Aid in Virtual Environments. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)*. ACM, New York, NY, USA, 267–274.

Nguyen-Thong Dang. 2007. A Survey and Classification of 3D Pointing Techniques. In *2007 IEEE International Conference on Research, Innovation and Vision for the Future*. 71–80.

Luigi Gallo, Giuseppe De Pietro, Antonio Coronato, and Ivana Marra. 2008. Toward a Natural Interface to Virtual Medical Imaging Environments. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '08)*. ACM, New York, NY, USA, 429–432.

S. Gebhardt, S. Pick, T. Oster, B. Hentschel, and T. Kuhlen. 2014. An evaluation of a smart-phone-based menu system for immersive virtual environments. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 31–34.

Tovi Grossman and Ravin Balakrishnan. 2006. The Design and Evaluation of Selection Techniques for 3D Volumetric Displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (UIST '06)*. ACM, New York, NY, USA, 3–12.

Cathleen E. Hughes, Lelin Zhang, Jurgen P. Schulze, Eve Edelstein, and Eduardo Macagno. 2013. CaveCAD: Architectural design in the CAVE. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. 193–194.

Jacek Jankowski. 2011. A Taskonomy of 3D Web Use. In *Proceedings of the 16th International Conference on 3D Web Technology (Web3D '11)*. ACM, New York, NY, USA, 93–100.

Beverley Laundry, Masood Masoodian, and Bill Rogers. 2010. Interaction with 3D Models on Large Displays Using 3D Input Techniques. In *Proceedings of the 11th International Conference of the NZ Chapter of the ACM Special Interest Group on Human-Computer Interaction (CHINZ '10)*. ACM, New York, NY, USA, 49–56.

P. Lubos, G. Bruder, and F. Steinicke. 2014. Analysis of direct selection in head-mounted display environments. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 11–18.

Mark Mine, Arun Yoganandan, and Dane Coffey. 2014. Making VR Work: Building a Real-world Immersive Modeling Application in the Virtual World. In *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction (SUI '14)*. ACM, New York, NY, USA, 80–89.

Michael Nitsche. 2008. Experiments in the Use of Game Technology for Pre-visualization. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share (Future Play '08)*. ACM, New York, NY, USA, 160–165.

Kevin Ponto, Ross Tredinnick, Aaron Bartholomew, Carrie Roy, Dan Szafir, Daniel Greenheck, and Joe Kohlmann. 2013. SculptUp: A rapid, immersive 3D modeling environment. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. 199–200.

Wolfgang Stuerzlinger and Robert J. Teather. 2014. Considerations for Targets in 3D Pointing Experiments. In *Proceedings of HCI Korea (HCIK '15)*. Hanbit Media, Inc., South Korea, 162–168.

Robert J. Teather and Wolfgang Stuerzlinger. 2014. Visual Aids in 3D Point Selection Experiments. In *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction (SUI '14)*. ACM, New York, NY, USA, 127–136.

Konrad Tollmar, David Demirdjian, and Trevor Darrell. 2004. Navigating in Virtual Environments Using a Vision-based Interface. In *Proceedings of the Third Nordic Conference on Human-computer Interaction (NordiCHI '04)*. ACM, New York, NY, USA, 113–120.

Jia Wang and Robert Lindeman. 2014. Coordinated 3D Interaction in Tablet- and HMD-based Hybrid Virtual Environments. In *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction (SUI '14)*. ACM, New York, NY, USA, 70–79.

Colin Ware and Steven Osborne. 1990. Exploration and Virtual Camera Control in Virtual Three Dimensional Environments. In *Proceedings of the 1990 Symposium on Interactive 3D Graphics (I3D '90)*. ACM, New York, NY, USA, 175–183.

Roy Sirui Yang, Alfonso Gastlum Strozzi, Anthony Lau, Christof Lutteroth, Yuk Hin Chan, and Patrice Delmas. 2012. Bimanual Natural User Interaction for 3D Modelling Application Using Stereo Computer Vision. In *Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction (CHINZ '12)*. ACM, New York, NY, USA, 44–51.

Ying Zhang, Reza Sotudeh, and Terrence Fernando. 2005. The Use of Visual and Auditory Feedback for Assembly Task Performance in a Virtual Environment. In *Proceedings of the 21st Spring Conference on Computer Graphics (SCCG '05)*. ACM, New York, NY, USA, 59–66.