Interaction With Partially Transparent Hands And Objects

Volkert Buchmann **Trond Nilsen** **Mark Billinghurst**

HIT Lab New Zealand
University of Canterbury
Christchurch, New Zealand,
Email: oakley.buchmann@hitlabnz.org

Abstract

Visual concealment of important objects and information by hands or tools can make many tasks more difficult. To alleviate this problem, hands and tools can be made partially transparent, allowing users to see their tools and work area simultaneously. This paper describes our experience with a system that can control the level of transparency of hands, tools and objects. We describe how users performed with uniform transparency across objects and with selective transparency where important details of objects are made less transparent. We identify several perceptual issues with this interface and propose possible solutions.

Keywords: Mediated reality, augmented reality, diminished reality, target acquisition, transparency

1 Introduction

When we interact with objects in our environment, our hands and tools are very close to the objects of interest and may cover relevant information. The occluded information might be critical, for example during surgery or when monitoring and operating controls such as in the cockpit of a plane. Some authors have proposed rendering hands partially transparent in Virtual Reality (VR) or Mediated Reality (MR) as a solution to these problems. However, no implementation exists yet, and we do not know how well users would perform.

Creating the illusion of transparency is an application of mediated reality (MR). As first described by Mann (1994) an MR system may augment reality with representations of virtual objects, hide representations of real objects or change the representation of real objects (Mann 1994). MR can be realised with a video see-through head-mounted display (HMD) displaying the environment as captured by one or two cameras mounted to the user's head. A computer in between the cameras and the HMD is able to manipulate the camera images before they are presented to the user.

A well known subset of MR is augmented reality (AR), where virtual objects are inserted into the real imagery in such a way that they appear to be part of the real environment (Azuma 1997). Another subset of MR is diminished reality (DR), where real objects are made invisible. For example, this is used to remove instruments from a surgeon's view (Mourgues, Devemay & Coste-Maniere 2001).

In order to assess potential usability issues of interaction with partially transparent hands and objects, we have implemented a simple MR system that uses video see-through to render the user's hand partially transparent. In this paper, we present initial results from this implementation. We describe how well people were able to use this interface for different tasks and identify a set of usability issues.

In the next section, we will summarize related work. In section 3, we will describe our implementation and how it performs.

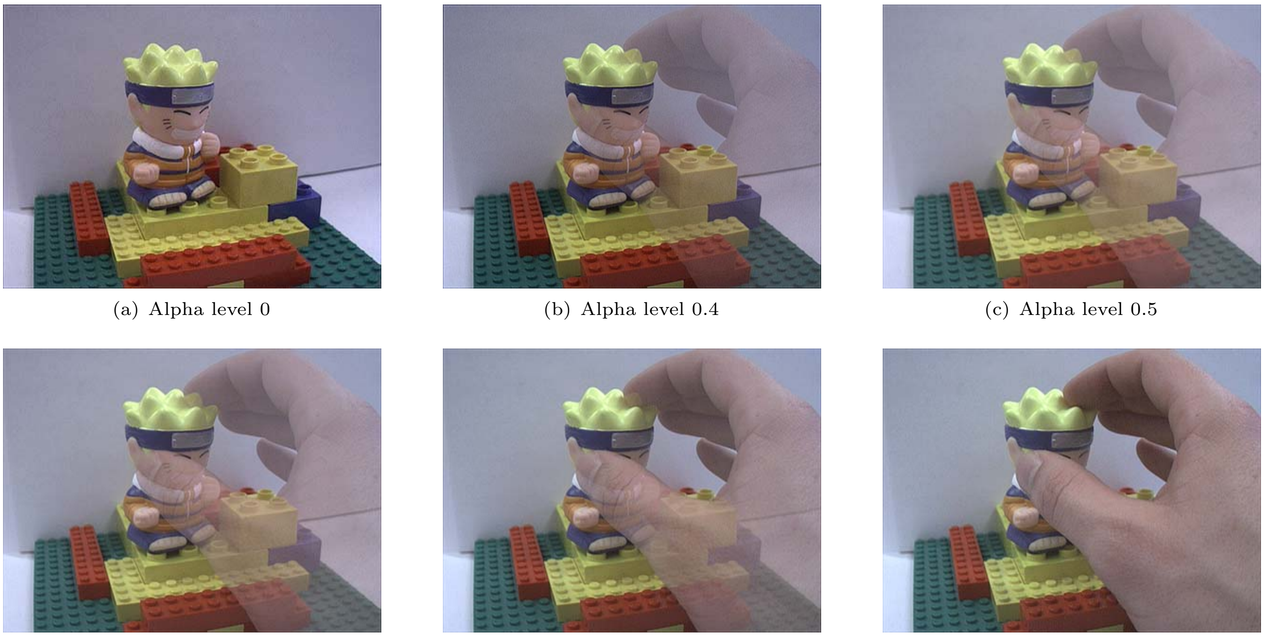
2 Related Work

The idea of making the user's hands partially transparent is not new; several authors have proposed this. However, there have been no implementations of a working system and no user testing on how such a system might affect performance.

Pierce, Forsberg, Conway, Hong, Zeleznik & Mine (1997) describe a gesture based object selection system for immersive 3D VR environments and note that the user's hands may occlude small or far away objects. They propose to render the virtual representations of the user's hands partially transparent to reveal these objects. Alternatively, they propose to render the hands more abstractly, for example as cursors.

Inami, Kawakami & Tachi (2003) describe a system to make real objects appear transparent. The objects that are to be made transparent are covered with retro-reflective material and a projector projects camera images of the objects' background onto them. To ensure that the projected picture is well aligned with the user's view, the image is projected via half-silvered mirrors attached to the user's glasses. The authors describe a non-wearable version, but suggest a head mounted projector could be used. The authors also suggest that this technique can be used to make surgeon's gloves transparent.

The two major implementation problems for partial transparency are the identification of the relevant objects in the current frame and the reconstruction of the obstructed background. A range of DR papers are concerned with hiding real objects from the user's view by replacing them with their estimated background. Most authors are concerned with making real objects completely transparent, while very few aim at partial transparency. Most authors are also concerned with hiding static objects such as buildings (Klinker, Stricker & Reiners 2001) or monuments (Lepetit & Berger 2001), but the possibility of removing instruments from a surgeon's view has also been explored (Mourgues et al. 2001). Methods used for generating the occluded background include additional cameras (Zokai, Esteve, Genc & Navab 2003, Kameda, Takemasa & Ohta 2004), dense temporal sequences of images as for example captured from a camera



(a) Alpha level 0
(b) Alpha level 0.4
(c) Alpha level 0.5
(d) Alpha level 0.6
(e) Alpha level 0.8
(f) Alpha level 1

Figure 1: Grasping an object with different levels of transparency of the hand.

(Lepetit & Berger 2001) and stereoscopic endoscopes (Mourgues et al. 2001).

To identify the user's hands in the current camera picture, optical tracking would be the most promising approach since it allows for exact registration with the image presented to the user. Possible options are for example the use of stereoscopic depth information (Gordan, Billinghurst, Bell, Woodfill, Kowalik, Erendi & Tilander 2002), recognising hands by their colour as provided by the functionality of the OpenCV toolkit[1] and the use of fiducial markers as used by the AR Toolkit[2].

3 Explorative Implementations

We implemented a simple system in order to find out how well partially transparent hands and objects may be used. Our implementation avoids the need for identifying hands and objects in the current camera frame and uses a very simple approach to background restoration. The camera is fixed and cannot be moved by the user, and the background must be static. At startup, the application stores a picture taken by the camera as the background image (figure 1(a)) and then blends each subsequent frame at the current alpha value against this background picture. For alpha-blending, we combine each pixel of the background image and the current camera frame using the standard alpha-blending formula. Objects that are present in the current camera frame but not in the background image are opaque at an alpha value of 1 and completely transparent at an alpha value of 0. Figure 1 shows the result of several alpha values. We found that it was easier for the user to view the generated image through an HMD at roughly the same position and pose as the camera rather than on a monitor. Although the viewpoint cannot be changed, this setup can provide valuable insights in how people use such a transparent interface.

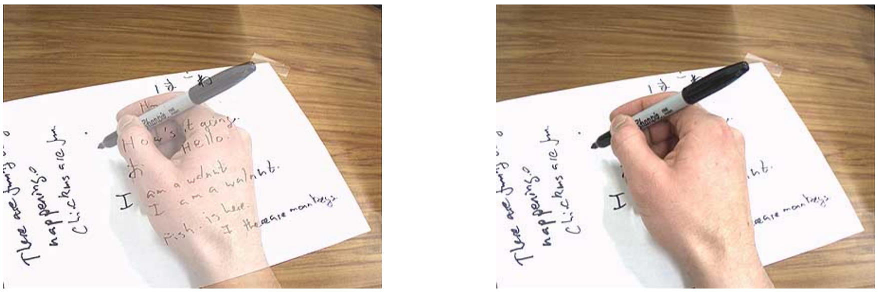
Our setup imposes several restrictions on what the user perceives. Stereoscopic depth cues are removed because we use only one camera. Cast shadows as depth cues are gradually removed by making them as transparent as the user's hand. We do not know if increased transparency of shadows decreases their efficiency, but we assume that making them less salient does have an effect; this is supported by the findings of Kersten, Mamassian & Knill (1997). In addition, our way of decoupling HMD and camera sometimes confuses the users because the user's view does not change according to their head movement and the camera angle might be wrong.

3.1 Uniform Transparency

In our first implementation, we rendered the reaching hand with alpha values 0.2, 0.4, 0.5, 0.6, 0.8 and 1 (see figure 1). We asked students from the HIT Lab for their opinion on the usability of the system for reaching towards and touching objects (figure 1), writing (figure 2), putting a screwdriver on a screw and inserting a USB card into a computer. Pen, screwdriver and USB card were rendered at the same transparency level as the user's hand. For each of these tasks, we started with an alpha value of 1 to let the users get accustomed to the task and then increased the level of transparency. When we noticed better performance with increased transparency, we suspected a learning effect and switched back to alpha value 1 for a comparison. Each session was performed in a relaxed atmosphere, no data was recorded automatically and we did not have questionnaires.

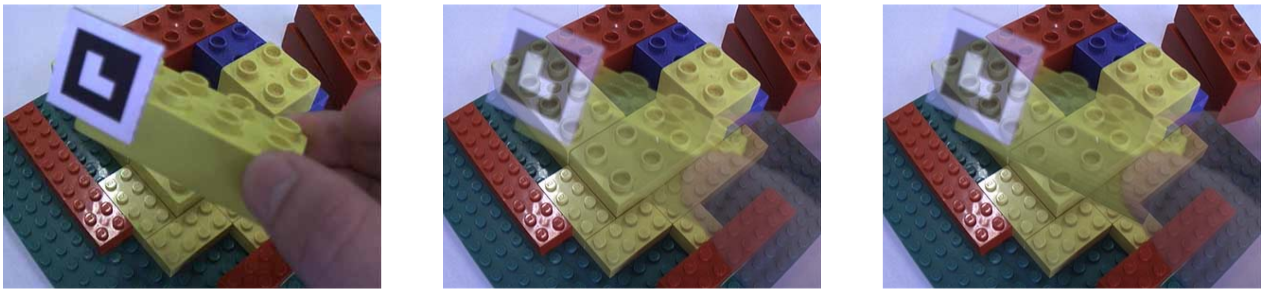
All users found the system fascinating to use and were certain that it would help them in situations were their hands or tools would cover relevant background structure. Students generally preferred alpha levels 0.6 and 0.8. They said that interaction was about as easy as with alpha level 1 while providing a "nice balance" between the hand and the background. Alpha levels 0.5 and below were regarded as too transparent for most tasks. Typical comments on alpha level 0.5 were "it feels a bit more alien", "kinda weird ... your mind is playing tricks on you" and "it feels like your hand is not really there" while levels 0.4 and 0.2 were completely dismissed. When comparing it to

[1] http://www.intel.com/research/mrl/research/opencv/
[2] http://www.hitl.washington.edu/artoolkit/



(a) Alpha level 0.5 (b) Alpha level 1

Figure 2: Writing with and without transparent rendering of hands. Note how the writing appears to be on the transparent hand rather than on the paper.



(a) Lego brick and hand are rendered opaquely. An AR Toolkit marker is attached to the brick so that its position and pose can be tracked.

(b) Lego brick and hand are rendered at a uniform alpha value of 0.6. At this level, it is difficult for the user to determine the exact position of the brick relative to the target brick.

(c) The bottom edge of the brick is rendered with an alpha value of 0.9. Now it is much easier for the user to place the brick.

Figure 3: Rendering the relevant edge of a Lego brick at very high alpha values improves interaction.

alpha level 0.5, one student said of alpha level 0.6 that it "feels more like I'm actually fiddling with the object". One student found that alpha level 0.8 was not transparent enough to reveal information about the background.

A general observation is that the perceived level of transparency varies depending on the background colour. If the background is darker, the hand will appear more transparent, while if the hand is darker, it appears less transparent (see for example figure 1(c)). This dependency of observed transparency is a potential problem for accurate perception of the user's hand. If this affects performance, transparency of the hand should adapt to the background light conditions, so that those parts of the hand covering darker regions are rendered with lower alpha values.

When asked, the users told us that they focussed on the edge regions of the hand, pen or USB card that were relevant for the current task. With decreasing alpha levels, the focus shifted from the edge regions to the actual edge since it became the only distinguishable feature. One user said that he got distracted by the background visible through his hand if he didn't concentrate on the USB card's edge.

We found two perceptual problems for alpha values 0.5 and lower. The first problem is that high transparency of the hand diminishes occlusion as a depth cue. Users reported that it seemed as if the hand might be behind the background objects, and that letters in the background seemed to be painted on the hand (figure 2(a)). See also figure 1 for how the index finger seems to disappear behind the figurine at low alpha levels. Occlusion is one of the most important depth cues (Cutting 1997) and highly relevant for hand based interaction in MR and AR (Buchmann, Violich, Billinghurst & Cockburn 2004). Bingham, Bradley, Bailey & Vinner (2001) found that elimination of occlusion has no effect when binocular vision is present. This suggests that stereoscopic vision might allow for lower alpha values of the hand. The second problem we discovered is that users felt that they had less control of their hand when it was very transparent, especially for precise movements. Mason, Walji, Lee & MacKenzie (2001) found that in an AR environment, reaching movement took longer when vision of the reaching limb was removed. By making the hand more and more transparent, we gradually removed high quality visual feedback of the hand, thus making it harder to control it.

3.2 Selective Transparency

From our experience with uniform transparency, we believe that an object can be treated as having regions of different task relevance. The bulk of the hand and object are not directly relevant for precise interaction, but for perceiving overall hand pose. Only those edges of the hand or tool that touch other objects are relevant for precise interaction. For example in the case of the USB card, users mainly concentrated on the edge of the card that is inserted into the slot. To actively support relevant edges, we used the AR Toolkit to track the position of a Lego brick (figure 3(a)) and rendered its bottom edge at very low transparency (figure 3(c)).

We hoped that this would allow greater levels of transparency of the object's main regions since the relevant edge of the brick was clearly visible. However, users still did not like low alpha values and again, 0.6 and 0.8 were picked as favourite levels. Overall, users liked the clearer view of the brick's edge that this interface provided and they were able to place the brick faster.

However, some noted that making the brick's edge less transparent also had drawbacks. With higher transparency of the brick's edge, it had been possible to see the otherwise occluded relevant part of the target brick. Future implementations should explore how thin such a low transparency region can be in order to provide better visibility of the occluded target structures.

4 Conclusion

We have presented an implementation of partial transparency for hands and tools in manual tasks, along with perceptual issues we observed from using our system. While users found that transparency helped them, they also saw drawbacks. At higher transparency levels, the lack of occlusion led to conflicting depth cues, particularly in our monoscopic setup. Further, the lack of visual feedback caused some users to report feeling reduced control over their hands. We implemented transparency as both uniform and selective according to task relevance of different regions of the hand. Users liked the decreased transparency of important regions in the selective rendering, but found that by this reduction of transparency, important background details were once again obscured.

Our initial results are promising and we intend to develop a system that works with a mobile camera rather than a fixed one. With this system, we will explore more visualisations such as adaptive transparency based on the brightness of the background, interaction with real and virtual objects and improving our selective transparency visualisation. Eventually, we intend to do a formal evaluation to determine the best interface configuration and to formally assess how well such a system performs.

Acknowledgement

We thank Kevin Sharp for suggesting us to render hands transparently.

We thank all students at the Hit Lab NZ who helped us with their ideas and suggestions.

This research was partially funded by Technology for Industry Fellowship grant TNNX0301.

References

Azuma, R. (1997), 'A survey of augmented reality', *Presence: Teleoperators and Virtual Environments* **6**(4), 355–385.

Bingham, G., Bradley, A., Bailey, M. & Vinner, R. (2001), 'Accommodation, occlusion and disparity matching are used to guide reaching: a comparison of actual versus virtual environments', *Journal of Experimental Psychology: Human Perception and Performance* **27**(6), 1314–1334.

Buchmann, V., Violich, S., Billinghurst, M. & Cockburn, A. (2004), Fingartips: gesture based direct manipulation in augmented reality, *in* 'Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia', ACM Press, pp. 212–221.

Cutting, J. E. (1997), 'How the eye measures reality and virtual reality', *Behavior Research Methods, Instruments & Computers* **29**(1), 27–36.

Gordan, G., Billinghurst, M., Bell, M., Woodfill, J., Kowalik, B., Erendi, A. & Tilander, J. (2002), The use of dense stereo range data in augmented reality., *in* 'Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002)', IEEE Press, pp. 14–23.

Inami, M., Kawakami, N. & Tachi, S. (2003), Optical camouflage using retro-reflective projection technology, *in* 'Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003.', pp. 348–349.

Kameda, Y., Takemasa, T. & Ohta, Y. (2004), Outdoor mixed reality utilizing surveillance cameras, *in* 'Proceedings of the 31st annual conference on Computer graphics and interactive techniques', ACM Press.

Kersten, D., Mamassian, P. & Knill, D. C. (1997), 'Moving cast shadows induce apparent motion in depth', *Perception* **26**, 171–192.

Klinker, G., Stricker, D. & Reiners, D. (2001), Augmented reality for exterior construction applications, *in* 'Fundamentals of Wearable Computers and Augmented Reality', pp. 379–427.

Lepetit, V. & Berger, M.-O. (2001), An intuitive tool for outlining objects in video sequences: Applications to augmented and diminished reality, *in* 'Proceedings of the second International Symposium on Mixed Reality, ISMR 2001', pp. 159–160.

Mann, S. (1994), Mediated reality, *in* 'Technical Report 260, M.I.T. Media Lab Perceptual Computing Section, Cambridge, Ma.'.

Mason, A. H., Walji, M. A., Lee, E. J. & MacKenzie, C. L. (2001), Reaching movements to augmented and graphic objects in virtual environments, *in* 'Proceedings of the SIGCHI conference on Human factors in computing systems', ACM Press, pp. 426–433.

Mourgues, F., Devemay, F. & Coste-Maniere, E. (2001), 3d reconstruction of the operating field for image overlay in 3d-endoscopic surgery, *in* 'IEEE and ACM International Symposium on Augmented Reality, 2001', pp. 191–192.

Pierce, J. S., Forsberg, A. S., Conway, M. J., Hong, S., Zeleznik, R. C. & Mine, M. R. (1997), Image plane interaction techniques in 3d immersive environments, *in* 'Proceedings of the 1997 symposium on Interactive 3D graphics', ACM Press, pp. 39–ff.

Zokai, S., Esteve, J., Genc, Y. & Navab, N. (2003), Multiview paraperspective projection model for diminished reality, *in* 'Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2003', pp. 217–226.