# User Evaluation of God-like Interaction Techniques

**Aaron Stafford** **Wayne Piekarski**

Wearable Computer Lab
School of Computer and Information Science
University of South Australia
Mawson Lakes Blvd, Mawson Lakes, South Australia 5095
Email: {aaron.stafford|wayne.piekarski}@unisa.edu.au

**Abstract**

God-like interaction is a metaphor for improved communication of situational and navigational information between outdoor users, equipped with mobile augmented reality systems, and indoor users, equipped with tabletop projector display systems. This paper presents the results of a user evaluation that explores user experience to a subset of the god-like interaction metaphor. The tasks performed by the participants were designed around known problems in the AR community. The results of the evaluation are intended to help define the boundaries in which the god-like interaction metaphor is practical for communication of navigational and situational information. This paper reports the findings of the evaluation as well as recommendations for further development of the interaction metaphor.

*Keywords:* Outdoor Augmented Reality, 3D reconstruction, user study, god-like interaction metaphor, remote communication.

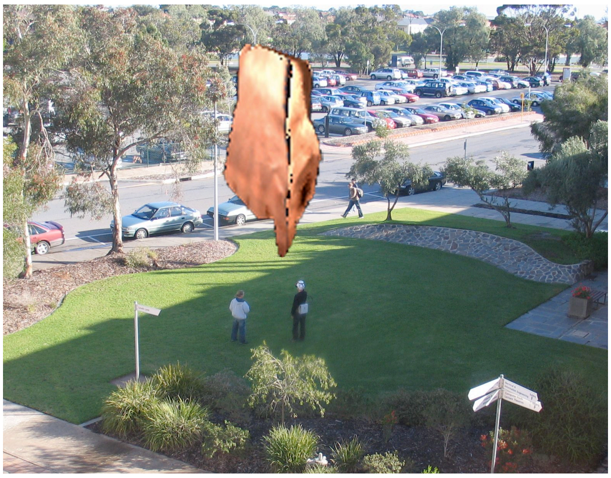

Figure 1: The area that the evaluation was performed in. The hand has been added to the image to indicate the area where the objects typically appeared and approximately what it looks like.

## 1 Introduction

Previously we presented god-like interaction (Stafford et al., 2006) as an interaction metaphor for improved communication between people located indoors, using a 3D reconstruction table system, and people located outdoors using an outdoor augmented reality (AR) system. The system enables users indoors to use their hands and physical props to provide navigational and situational information to users working remotely outdoors. Users' hands and physical props are captured by a number of cameras around the edge of the table. This imagery is sent to the person outside. Using an outdoor AR system they see the imagery anchored to a real world location (see Figure 1). The capabilities of this system enable a wide range of intuitive interactions to be used to communicate situational and navigation information, some more helpful than others.

Initial studies are required to quantify how understandable the visual information is to the outside user. This paper presents an evaluation of a subset of the god-like interaction metaphor. The goal of this evaluation is to improve our understanding of user comprehension of imagery and intentions presented in an outdoor AR context. The tasks of the study are based on known difficulties in the AR field, such as depth perception and 3D reconstruction from a small number of view points, to gauge the impact of the limitations on our metaphor. The limitations can then be addressed to improve overall acceptability of this interaction metaphor. This evaluation is primarily qualitative in nature as participant feedback and responses to open ended questions provide the source of most of the results.

Outdoor AR studies are few in number (Thomas et al., 2002; Cheok et al., 2004; Wither and Höllerer, 2004, 2005; Avery et al., 2006). They are technically challenging. Without quality hardware systems it is difficult to accurately determine user location and orientation, two elements critical for rendering realistic augmented scenes. If these technical challenges are not addressed the user experience is poor. Systems for running outdoor AR applications require a significant amount of complex hardware but also need to be light enough so that users can move freely without discomfort

This paper presents an overview of god-like interaction and studies in outdoor AR in Section 2. The evaluation that we undertook is presented in Section 3 with the results in Section 4. A conclusion and future work is presented in Section 5.

## 2 Background

Computer supported cooperative work in the form of remote sensing has been around from some time. With these systems remote workers can be instructed to perform tasks by an expert in situations where it is impractical for the expert to be physically present at the site of interest. SharedView (Kuzuoka, 1992) is a system that attempts to convey social interaction cues such as gestures. The remote worker wears a camera mounted on a helmet so that the expert can see what the remote worker sees. The expert can point to items of interest on a display, the gestures are captured by another camera and sent to the remote worker. The gestures are displayed to the remote worker on a head mounted display (HMD) such that they can effectively see what the expert is pointing at. This system was shown to be more effective that using voice alone.

A similar system, Netman (Bauer et al., 1998) was later developed that also incorporated an HMD and camera worn by a remote user such that an expert could monitor their work from a distant location. This system also enables remote users to interact with a database to extract information about the area of interest. Experts can push documents to the remote worker that are of relevance to create a conversational context.

The MARS test bed (Höllerer et al., 1999) enabled remote workers to work in an outdoor environment. The system also has an emphasis on interaction with information that is spatially registered. Indoor experts can work with a 3D model to generate content that is registered to GPS coordinates. By tracking the remote workers head movements with an orientation sensor and monitoring their position with a GPS system the spatial information appears to be anchored to real world locations.

Of increasing focus is the role of the display devices used by the expert in the remote sensing scenario. Large displays are frequently used in a horizontal orientation to give the expert a bird's eye view of the situation in which the remote workers are in (Krum et al., 2002; Kurata et al., 2005). This situational visualisation research has shifted the focus from the interaction between the remote worker and the expert to the interaction between the expert and the visualisation system.

The goal of our god-like interaction is to present users with a metaphor for interacting with our system rather than a detailed list of many specific ways that it can be operated. By explaining the overall concepts to users, such as how they have assumed a god-like role and that outdoor people can see and hear everything occurring on the table, we believe users should be able to operate the system intuitively.

We have previously constructed a system (Stafford et al., 2006) to enable us to explore this interaction metaphor and to evaluate the effectiveness of the metaphor for remote communication. This system previously enabled us to trial a number of prototype applications including:

**Navigation:** Hand gestures such as pointing are a simple way for people indoors to provide navigational or situational information to people working outdoors. Indoor users can simply point to a location on a map and speak a verbal command. The indoor user's hand is rendered at the corresponding physical location for users outdoors. Indoor users can also point and drag their fingers to indicate a path to follow or a boundary to avoid crossing.

**Augmented Post-It notes:** Indoors users can write information on Post-It notes and stick them to the surface of props and place them on the table. The quality of the reconstruction is such that hand written notes are legible to people outside. An important part of this technique is that no input devices such as keyboards or mice are required to implement this functionality, and instead standard office supplies are used as in daily life.

**Relocation:** Allows the system to be able to operate over a large region even though the working volume for object capture might be quite small. Even if the working volume could be expanded to the entire size of the table, the relocation mechanism is still useful for when indoor users need to work over a very large outdoor region, where scaling the needed outdoor region to the size of the table would make it impractical to view features on the map.

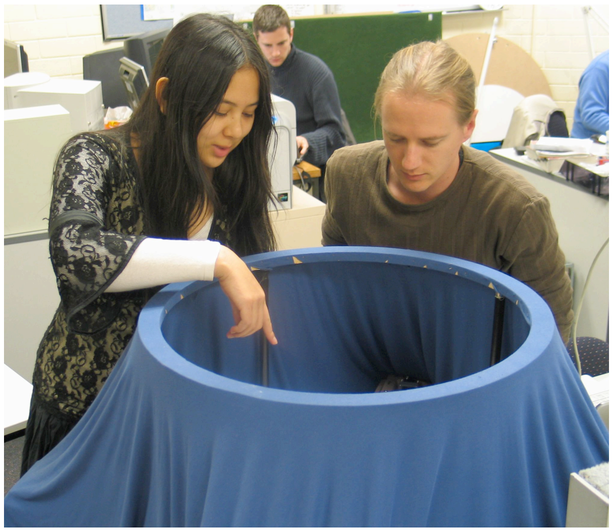

Figure 2: Tabletop display system: Two indoor users providing navigational information to users outdoors.

**Snapshot:** The snapshot feature enables us to populate a large area with many props even though we can only use one prop at a time on the table. The snapshot feature takes the current reconstruction and makes a copy of all the data needed to reconstruct it in the future, effectively pausing the copy so that it appears frozen.

We use Li's Video-Based Rendering algorithm (VBR) (Li et al., 2003) to generate a texture mapped visual hull (Laurentini, 1994) of objects placed onto the table. It is necessary to separate users' gestures from the rest of their bodies, consequently the tabletop device resembles a miniature movie set, as can be seen in Figure 2. The table has a blue-coloured wall perimeter, so that props placed onto the table are easily segmented from the background. A projector mounted above renders a top-down view of the remote users' working area onto the table surface. The cameras are recessed into the surface of the table so that they are able to only see objects above the plane of the table surface and unable to see the projected image.

We use our Tinmith mobile outdoor AR system (Piekarski and Thomas, 2003*b*) to provide outdoor users with an AR view of the world. The existing modelling software is used as-is, with extra support added to support the decoding and rendering of real-time captured objects. We connect the mobile backpack system with the indoor tabletop system via an 802.11g wireless network. Using the modelling techniques built into Tinmith (Piekarski and Thomas, 2003*a*), outdoor users can select the reconstructed objects and manipulated them. By using two hands a user can rotate or scale a reconstruction. Relocation is performed by slaving the location of the reconstruction to a 3D cursor on the gloves, pinching gestures move the reconstruction closer or further away.

Outdoor AR gaming is currently the richest source for studies in the field of outdoor AR. Thomas et al. (2002) studied the usability and playability of ARQuake, an outdoor AR adaptation of the popular desktop game Quake. In this study the game was played at a number of discreet locations along Milgram's Reality-Virtuality continuum (Milgram and

Kishino, 1994). Registration between the real world and the game world was found to be a significant problem due to the mismatch between the small field of view of the HMD and the relatively large field of view of the game world. Registration mismatch caused objects to appear closer than they actually were and some participants described the experience as uncomfortable. The study also found that virtual shadows (shadows from virtual object projected onto the real world) detracted from game-play experience as they were found to be "annoying".

More recently virtual shadows have been shown to improve the feeling presence of a virtual object (Sugano et al., 2003). The authors also showed that accurate shadowing improves three dimensional perception of the scene. While these experiments were performed indoors with controlled lighting it is likely that similar effects would be observed outdoors given accurately generated shadows. However rendering shadows accurately outdoors is harder due to constantly changing lighting conditions. So while more improvements to rendering shadows of virtual objects have been made (Jacobs et al., 2005; Haller et al., 2003; Kakuta et al., 2005) to date no one has explored the use of shadow generation for virtual objects outdoors.

Human Pacman (Cheok et al., 2004) is the first example of a collaborative outdoor AR game. It is an outdoor AR adaptation of the arcade game where players assume the role of Pacman, a ghost or a "helper". The authors conducted a study comparing the enjoyment of the original Pacman to the outdoor AR version they developed. Results show that participants found Human Pacman significantly favourable to normal Pacman. Sensor drifting was prevalent and the authors suggest that it reduced the overall enjoyment experienced by the participants.

Avery et al. (2006) conducted a user study to measure user enjoyment of Sky Invaders 3D, an outdoor AR game based on the arcade game Space Invaders. Their study compared the playability of Sky Invaders 3D with a desktop version of the same game. They found that the outdoor AR version of the game was significantly more enjoyable and more intuitive to use compared to the desktop version. The study restricts its findings to the context of outdoor AR gaming.

Beyond outdoor AR gaming Wither and Höllerer (2004) compared different input devices for interacting at a distance in outdoor AR. They compared a Twiddler2 keyboard, a RocketMouse finger trackball, head orientation with two buttons and head orientation alone. These devices were compared to determine the best device for manipulating a 3D cursor. Their results show that the RocketMouse finger trackball and the head orientation with two buttons were equally the fastest. Interestingly, study participants thought the RocketMouse finger trackball was significantly faster than the head orientation with two buttons.

Wither and Höllerer (2005) later conducted a study comparing several techniques for providing depth cue for 3D cursor placement in an outdoor AR context. They found that colour encoded depth markers reduced distance estimation errors approximately 50%. A top-down view was the most preferred cue for estimating distance however along with shadow planes they were not significantly helpful to users for refining distance estimations.

## 3 Evaluation

The evaluation took place outside the main building of the School of Computer and Information Science in a grassy area seen in Figure 1. Participants wore the Tinmith system with a video see-through HMD for the entire experiment. The HMD uses I-glasses SVGA 3D PRO with an SVGA video resolution of 800 x 600, however the HMD is feed via S-Video so the effective resolution is actually 720 x 576 and the field of view is 26 degrees. The video to the HMD is captured by a Point Grey DragonFly camera.

Figure 3: The six reconstructions shown to the evaluation participants: a small road sign, a pointing hand, a good luck cat, a can of coke, a flavoured milk carton, and a jar of Vegemite.

Participants were asked to complete four outdoor tasks (described below) and a questionnaire. At the start of each task participants were given some information about the task to be performed. For each task participants were shown a number of reconstructions and asked a series of questions which were recorded by the experimenter. Other observations made during the evaluation were also recorded. After the outdoor tasks were completed the participants filled in a short questionnaire about what they had just experienced.

The reconstructions seen by the participants were all captured using the reconstruction system described in (Stafford et al., 2006). The cameras used by the reconstruction system are Point Grey DragonFlys capturing at 640 x 240. For this evaluation, all objects and animations were captured, processed and loaded onto the Tinmith system offline. This eliminated the need for a second participant operating the table for each outdoor participant, reducing the complexity of the evaluation and ensuring that all outdoor participants shared the same experience. The remainder of this section describes the four outdoor tasks and the questionnaire.

### 3.1 Recognition

The goal of the recognition task was to determine if objects constructed using information from only four cameras was sufficient for participants to easily identify the object. The objects we used were assorted props found in and around the lab. Each of the reconstructions in this task were essentially snap-shotted, that is, they are frozen in time. During capture each object was placed facing a particular camera to obtain the best possible reconstruction. Objects not facing the cameras are often completely unrecognisable which is primarily an artefact of the small number of cameras.

During the evaluation participants were sequentially shown six reconstructions (as seen in Figure 3). The reconstructions were displayed on a grassy area approximately 3 metres from participants' initial starting position (as seen in Figure 1). For each reconstruction, participants were asked what they thought the object was. Responses were recorded by the experimenter. The experimenter controlled the rate at which objects were displayed via a Bluetooth controller, this meant that participants could walk around or spend time looking at a particular reconstruction if they wanted to. At the end of the task participants were asked if they had any comments about what they had just seen, and again responses were recorded by the experimenter.

### 3.2 Interpretation of Intent

The goal of the interpretation of intent task was to determine if participants would be able to infer meaning from a short animation of a hand performing an action. All animations where shown without any corresponding audio. The first animation was of a hand coming out of the sky pointing to a particular location - the intended meaning of this action was either "go to this location" or "there is something interesting at this location". The second animation, as seen in Figure 4, was of a fist with the index finger pointing out and shaking side to side, like when a parent scolds a young child. It is intended to be recognisable as a negative, "no", "don't do that" type of action. The third gesture was a thumbs up gesture, as seen in Figure 4, intended to convey an affirmative meaning. For each animation, participants were asked to describe what information they thought the hand was trying to convey to them. Answers were recorded by the experimenter.

### 3.3 Judging Distance

The goal of this task was to determine how technology is used for determining distances in an outdoor AR setting and if users could consistently judge the distance of the pointing hand when rendered at various sizes. Participants were shown the pointing hand at one of two distances (as seen in Figure 5). The first was a location 50 metres away from participants' GPS location. The second was a location 100 metres from participants' GPS location. The hand remained anchored to a location until the next hand was rendered. At each one of these locations three hand sizes were rendered: 2 metres tall, 5 metres tall or 10 metres tall. Therefore participants were asked to determine the distance of a total 6 hands. Participants were not told that different hand sizes would be used. The order in which the 6 hands were rendered was previously determined and each participants saw the hands appear in the same order.

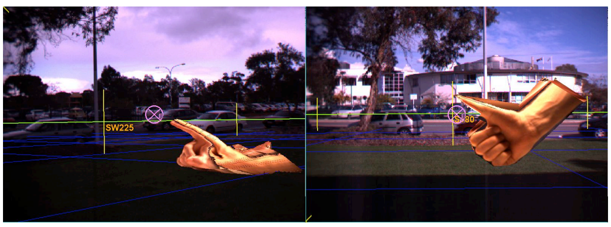


Figure 4: A screen capture of the no, no, no gesture is shown on the left and on the right is a screen capture of the thumbs up gesture.


Figure 5: The 6 hands in order of distance from participants (as perceived by the participants) they are (left to right, top to bottom): largest at 50 metres, medium at 50 metres, largest at 100 metres, medium at 100 metres, smallest at 50 metres and smallest at 100 metres.

Participants were not told the two distances the hand would be rendered at. When each hand was shown participants were asked to determine the distance in metres. The experimenter noted whether or not participants moved from their initial location whilst determining distance to see if they were relying on any motion cues.

### 3.4 Locating Objects

The goal of this task was to determine user ability to locate the hand without any other cues.

The pointing hand was rendered in the AR world but not in the immediate view of the participant, so it forced them to look around to locate the hand. The hand was either located in line with the horizon or at 45 degrees above the horizon. It was rendered at a location either 45 degrees to their right, 45 degrees to their left or behind them. Participants were asked to locate the hand a total of 12 times. In 3 cases the hand was not being rendered. In 3 cases it was being rendered at eye level and in 6 cases it was being rendered at 45 degrees up. Some screen captures taken from the locating tasks can been seen in Figure 6.

Participants were asked to verbally indicate they had located the hand. Alternatively if they believed that the hand was not being displayed they were asked to say this was the case. For each hand, the participants were timed to determine how long it took to locate, or work out that it was not being rendered.

### 3.5 Comfort

The final task of the evaluation was a questionnaire regarding the wearability of the current outdoor AR system. The current outdoor AR system we used is mounted on an army style padded belt. The HMD, orientation sensor, GPS and video camera are mounted to a skating helmet and weighs 1.5kg. The belt holds a 4kg CPU box and a 2kg battery. Participants were asked to complete a questionnaire at the


Figure 6: Screen captures from the locating tasks. Clockwise from top left: the hand amongst some trees, the hand against a plain sky with good contrast, the hand against a sky of low contrast and the hand in front of a building with a similar colour.

conclusion of the evaluation to determine the physical demand the belt and helmet placed on their body. The following questions were asked on a 5-point Likert scale:

- How aware of the technology were you?
- How aware of the weight of the technology were you?
- Did you find the field of view of the HMD adequate for the task?
- Did you find the backpack comfortable to wear?

Fourteen parts of the body were identified: neck, hands, wrists, fingers, elbows, arms, shoulders, back, waist, legs, knees, ankles, feet and eyes. For each part participants were asked to rate the level of discomfort experienced as a result of wearing the system for the duration of the evaluation. The questionnaire also had room for participants to specify any other parts of the body they believed were affected by the system.

## 4 Results

Twenty-one participants made up of staff and students from the School of Computer and Information Science took part in this evaluation. There were 17 males and 4 females. Just over half of the participants had had little or no experience with AR. However the group were highly computer literate with users claiming an average 45.95 hours a week spent interacting with desktop computers.

### 4.1 Recognition

Participants were generally able to correctly identify the objects (see Table 1) however comments on the quality of the reconstruction were common.

| **Object** | **Percent Correct** |
|---|---|
| Street sign | 100% |
| Hand Pointing | 100% |
| Toy cat | 100% |
| Coke can | 76.2% |
| Flavoured milk carton | 95.2% |
| Vegemite | 95.2% |

Table 1: Summary of participants' ability to recognise objects.

Many participants did not correctly identify the coke can. Several participants suggested that it was some sort of rectangular prism that had been textured with coke-like textures. One participant suggested that the coke can looked like a coke phone booth, another that it looked like a coke machine, yet another thought that it was a garbage can with coke signs on it.

Due to the reconstruction the normally cylindrical can has taken a box shape. Participants' comments suggest that some were using shape and size as a cue for determining the object. Reconstructions that are square, such as the carton of milk, were considered to be faithful representations of the original object. This coincides with the most common comment which was that objects looked square. The shape is a direct result of the number of cameras used to perform the reconstruction. The current reconstruction system uses four cameras placed at equal distances around the object to be reconstructed. Increasing the number of cameras will generate a reconstruction that is closer to the original shape of the object. Increasing the number of cameras will have two clear benefits: objects will take a shape closer to their original form and objects will not need to be as carefully placed in the system in order for them to look reasonable.

Several participants pointed out that the objects did not fit in with the background since the sun does not light it up like it does with objects in the real world. The same group of participants also commented that the reconstructed objects didn't cast shadows.

Results from this task indicated that the majority of objects were clearly recognisable. This is an encouraging result which means that common objects located within the workplace make good props for the god-like interaction metaphor because they are easily identifiable even though not perfectly reconstructed.

### 4.2 Interpretation of Intent

The results of this task show that 100% of participants identified that it was a hand pointing down towards the ground. When asked what they thought the intention of the action was the response was varied, but the general consensus was that the hand was pointing to something interesting and that the intended meaning was to go and have a look at it. The thumbs up was interpreted as an affirmation 100% of the time and the finger waving gesture was interpretive as a negative gesture 95.2% of the time with only one person not understanding the intended meaning.

The results show that participants working indoors can provide yes/no responses using just gestures and without using audio. Performing a gesture indicating an action appears to be more difficult as the gestures seemed to be more open to interpretation than the yes/no type gestures. Therefore it is necessary to provide another cue such as audio to provide extra information while performing complex gestures.

Positive and negative cues could be coupled with an action in order to provide information about tasks to be performed. Instructions can be given as a concatenation of a series of gestures, for example, action + positive or negative affirmation. One example is: "go here" (positive affirmation), "go here" (negative affirmation) - implied meaning: "go here, but don't go here".

### 4.3 Judging Distance

Figure 5 shows the 6 hands in order of closest to farthest according to participants' responses. At an actual distance of 100m the smallest hand was guessed to be at 28m, the medium hand at 23m and the largest hand at 21m. At an actual distance of 50m the smallest hand was guessed to be at 28m, the medium hand at 17m and the largest hand at 13m. In order to assist estimations, 18.2% of guesses involved a participant moving from their initial location. On average this was no more that 2 metres.

#### 4.3.1 Size

The data indicates that participants predominantly used size to determine distance. Looking at the order that participants placed the hands (see Figure 5) there is a clear progression from largest to smallest. This is emphasised by the fact that participants incorrectly guessed hand five (in Figure 5) to be further away than two other hands that were actually further away.

These results can be explained using depth estimation studies such as those by Cutting (2002). Cutting shows how depth perception is formed using nine cues which are detected in a variety of ways. Having prior knowledge of the size of an object is an important cue needed to estimate the size of the object, and without it this affects estimation accuracy, particularly at far distances. Cutting describes how perception can be divided into three regions, 0-2 metres, 2-30 metres, and 30 onwards. From 30 metres onwards, knowledge of the size of the object is one of the only cues available, and so this explains why the estimation was poor when the hand was located at 50 or 100 metres.

The implications for god-like interaction is that it is not a good idea to adjust the scale of the hand during interaction as it is likely to lead to confusion about how far away it is. Users will apply their previous experience and guess objects to be closer or further away depending on the scaling applied. Scaling by the outdoor user is unlikely to affect their perception.

Distance estimation was poor in general, for example, with the hand at 100 metres participants were on average 75 metres short. Since accurately determining distance would be useful for god-like interaction in the future a range counter will be associated with each object in the AR world to provide instantaneous and accurate description of distance. However this is unlikely to improve the perceived distance, as the hand would not look any different.

#### 4.3.2 Occlusion

Occlusion is a significant problem for outdoor AR systems. Many participants commented that the hand was being rendered in front of an object, such as a parked car, that was closer than what they thought the hand was. But because the hand was rendered in front of the object due to the video overlay they felt it must actually be closer than the object. Therefore participants often guessed the hand to be slightly closer than the closest occluded object.

Livingston et al. (2003) conducted a user study to determine the best approach for rendering occlusions. Results indicated that a combination of wire+fill with a decreasing opacity and decreasing intensity is as good as, if not better than, using a ground plane. Although the participants of the study were seated indoors, the results are likely to be applicable to occlusions in an outdoor AR context. We plan to use a decreasing opacity technique such that as the distance between the observer and the hand increases the opacity decreases. We will then test whether this technique reduces the effect associated with occluded objects.

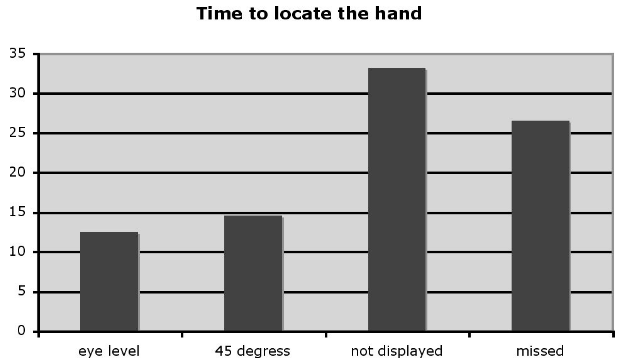


Figure 7: Average times to locate the hand at: eye level, 45 degrees, not displayed, incorrectly identified as not being displayed.

#### 4.3.3 Height

A number of participants indicated that the hand was pointing down to a car. Therefore they thought that the hand must be at about the same distance as the car. This is not a case of occlusion as such but more so that the participants related the distance of the hand to a physical object that they thought the hand might actually be pointing to. This has serious implications for the god-like interaction metaphor because it is hard for users indoors to know if the hand seems to be pointing to a physical object in between the location that they are actually pointing to and the location of the participant. Again this is a difficult problem and it would be interesting to try fading the hand the further away it gets so that users can attempt to guess the distance based on the opacity of the hand.

Participants used the height of the object sometimes to guess the distances - participants indicated that if an object was higher they thought it was further away. This appears to be supported by the results. The second and third hand in Figure 4 look to be the same size, the only noticeable difference is that the height of the two hands appears to be different with respect to the AR ground plane. However participants still managed to put the two hands in the correct order. The second hand is closer to the ground plane than the third hand indicating that participants resolved the difference using the height information.

### 4.4 Locating Objects

On average it took participants 12.4 seconds to locate an object at eye level, 14.5 seconds to locate an object at 45 degrees up and 33.1 seconds to determine that the hand was not being displayed (see Figure 7). The hands that were mistakenly identified as not present (6%) were determined as such on average in 26.4 seconds. Since the average time to correctly determine that a hand was not present was nearly 7 seconds more, it is likely that the reason that these objects were missed was due to insufficient time spent looking for the objects and therefore the objects probably never came into the field of view.

91.6% of objects missed were at 45 degrees. In fact there was only one object at eye level that was missed.

The hand located 45 degrees above the horizon took more time to locate than objects at eye level. This can be attributed to the vertical field of view of the camera mounted on the HMD which is 23 degrees. Therefore a scan of the horizon would not reveal objects that might otherwise be visible if no HMD was being worn. Additionally participants generally scanned the horizon before searching the sky.

Participants commented that if the background had similar colours to the hand then it was harder to locate the object. Conversely participants indicated that it was easier to locate objects if the contrast between the hand and the background image was high. The factors that affect the contrast can be broken into two categories:

1. Camera related
    - The camera has a number of properties and settings that can affect the contrast of the hand against the real world background. This includes exposure time, brightness, and white balance. The most noticeable effect is caused by insufficient ambient light; in this case significant colour loss is experienced.
2. Real world related
    - If participants are looking at a clear sky then the hand becomes easy to locate, as the background is mostly planar.
    - If the background is colourful then it is more difficult to locate the object
    - During the evening when there is less light the contrast is similar to the previously mentioned camera effect where the background becomes rather dark. This implies that deliberately increasing the contrast by darkening the background should lead to improved times to locate the hand.

While the goal of this task was to determine ability to locate objects not in immediate view without other cues it is worth mentioning that other cues, such as steering arrows that point to off screen objects, could be useful for assisting users in locating the hand. However as the number of reconstructions increases, distinguishing between arrows would become increasingly difficult. Alternatively attention funnels (Biocca et al., 2006) have been shown to be very effective for gaining users attention and guiding users to virtual or real objects that have predefined spatial coordinates assigned to them. It is possible that attention funnels could be used to direct attention to reconstructions produced by our system as each reconstruction has spatial coordinates defined in GPS coordinates.

### 4.5 Comfort

The average response to the question, “How aware of the technology were you?” was 3.9 (see Figure 8), where 1 was “not at all” and 5 was “significantly”. This question was aimed at identifying how conscious people were of the fact that they were wearing the outdoor AR system. Although the response to this question indicates participants’ obvious awareness of the system we predict that as the weight of future backpack systems decrease, so will awareness of the system. Therefore this result will be useful for future comparisons.

The average response to the question: “How aware of the weight of the technology were you”, was 3.0 (see Figure 9), where 1 was “not at all” and 5 was “significantly”. This result suggests that overall awareness of the technology was not predominantly caused by the weight of the system and is therefore likely to be more to do with the HMD. Future studies will breakdown the awareness further to include participants’ responses to awareness of the HMD.

The average response to the question: “Did you find the field of view of the HMD adequate for the task” was 3.2 (see Figure 10), where 1 was “not at all” and 5 was “adequate”. This was higher than expected and is an encouraging result as it suggests that many participants found the field of view adequate for the task. It would be useful in future to determine the adequacy of the field of view for individual tasks.

The average response to the question: “Did you find the back pack comfortable to wear” was 3.7 (see Figure 11), where 1 was “not at all” and 5 was “comfortable”. Our results show that 25% of participants found the belt designed pack back comfortable. This is comparable to the result found by Cheok et al. (2004) for their outdoor AR system.

There were four areas of the body that participants commonly identified as having obvious discomfort due to wearing the outdoor AR system: The neck, the waist, the eyes and the back. The neck had an average response of 2.4 (see Figure 12). During the locating objects task participants repeatedly looked up 45 degrees from the horizon in order to located the hand. As mentioned in Section 3.5 the weight of the helmet is 1.5kg combined with the actions performed by the participants it is not surprising that the neck

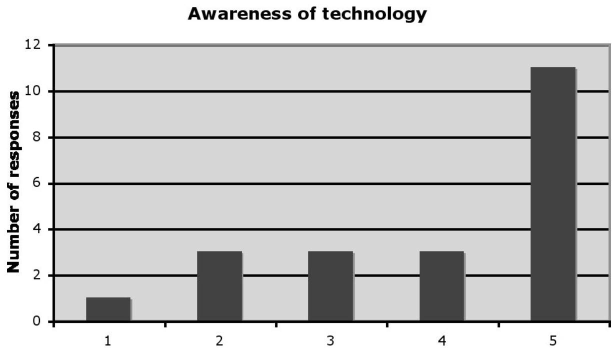


Figure 8: Participants’ responses to awareness of technology

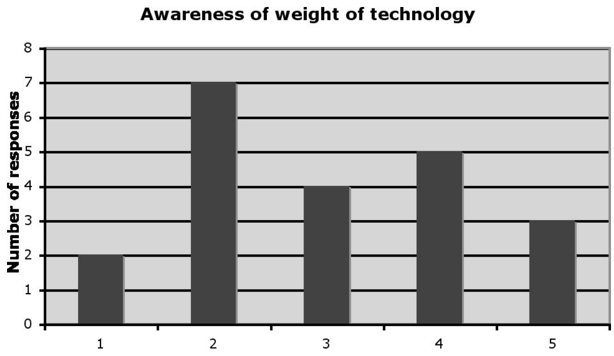


Figure 9: Participants’ responses to awareness of weight of technology

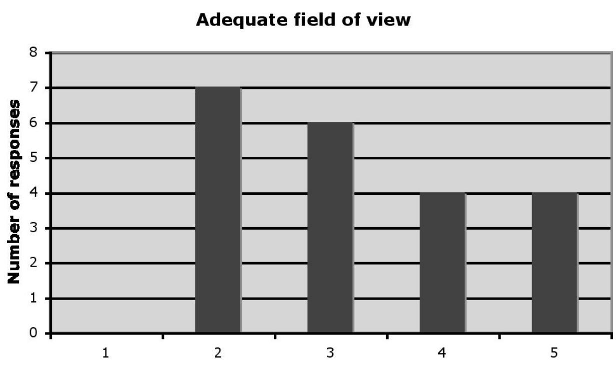


Figure 10: Participants’ responses to adequate field of view

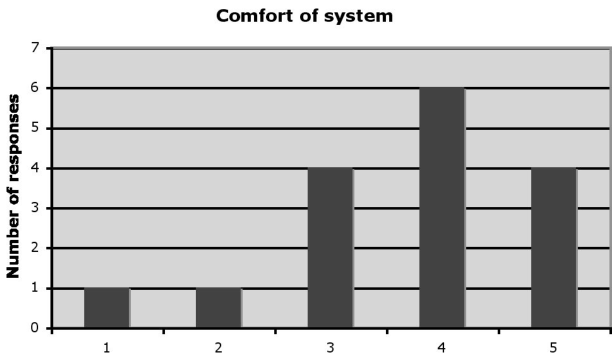


Figure 11: Participants' responses to comfort of the system

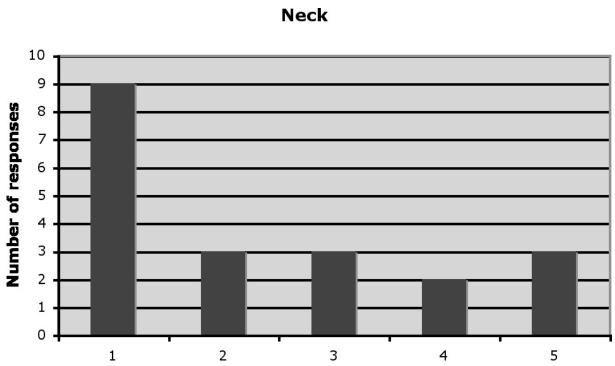


Figure 12: Participants' responses to discomfort in the neck.

was the region with equal the most discomfort.

Participants also marked the waist as having the same level of discomfort with an average response of 2.4 (see Figure 13). As the waist is the region carrying the most weight this result is understandable.

Discomfort in the eyes had an average response of 2.3 (see Figure 14) this is likely to be a combination of two factors. The first is some users' sensitivity to the display causing fatigue in the eye. The second is the design of the helmet, participants who rated discomfort in the eyes the highest also made comments such as "my forehead was slightly compressed", "there was strong pressure on my forehead". As the eyes are the only part of the head that were on the questionnaire it is possible that they received some high scores for areas around the eye rather than the eye itself. Future analysis of our backpack designs will need to breakdown the discomfort experienced on the head into smaller parts to better understand discomfort experienced.

Discomfort in the back had an average response

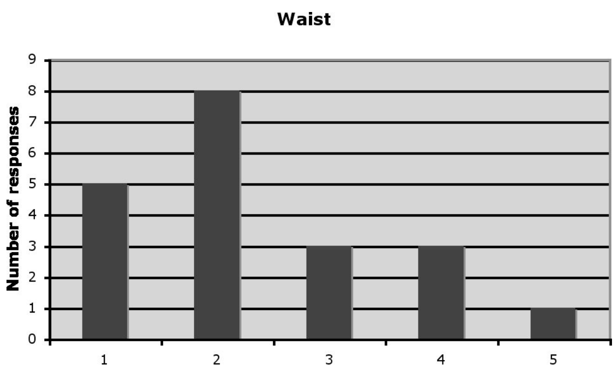


Figure 13: Participants' responses to discomfort experienced in the waist.

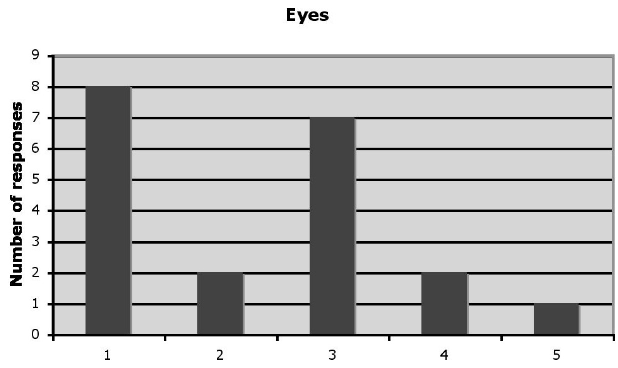


Figure 14: Participants' responses to discomfort experienced in the eyes.

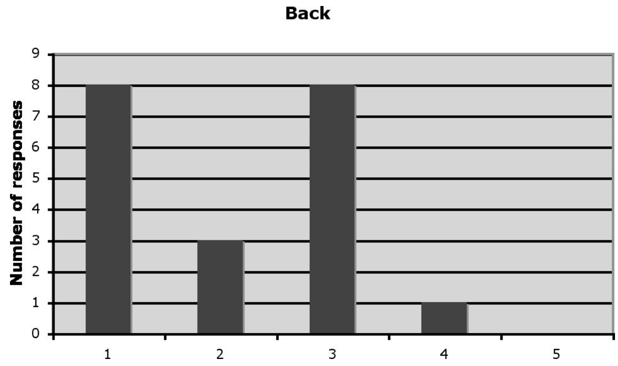


Figure 15: Participants' responses to discomfort experienced in the back.

of 2.1 (see Figure 15). Given the weight of the technology and the discomfort experience in participants' waists and necks it is not surprising that a relatively high level of discomfort was also experienced in participants' backs.

## 5 Conclusion and Future Work

There are a number of well known limitations in the AR research field. It is important to understand the effect these limitations have on our god-like interaction metaphor. Depth perception is an important cue for understanding the environment. Using the pointing hand as a navigational aid will need to be considered further as judging distance is an obvious problem. Scaling of the hand by a user working indoors during interaction should be avoided as it leads to confusion about perceived distance. Adding a distance counter and making the hand more transparent the further it is away from a user is likely to assist in reducing the effects of occlusion.

The quality of a reconstructed visual hull is typically limited by the number of cameras used to perform the reconstruction. The research in this paper has shown that with only 4 cameras a reconstruction is typically good enough to be identified by an outdoor user provided they are carefully place. Future enhancements of the tabletop reconstruction system will focus on increasing the number of cameras so that physical props can be placed less formally.

The quality of the reconstructions also appear to be sufficient for outdoor users to interpret simple yes/no type hand gestures. However as the pointing hand was interpreted to mean a number of different things it is clear that other cues such as audio are necessary to convey intended meaning of more complex gestures.

This paper presents one of the few studies conducted using an outdoor AR system. It is important to test well know AR limitations in an outdoor AR setting to determine if certain assumptions hold true. This evaluation has highlighted practical information about outdoor user experience of a subset of the god-like interaction metaphor. Results of this study will be used to make improvements to the current system.

## References


Avery, B., Piekarski, W., Warren, J. and Thomas, B. H. (2006), Evaluation of user satisfaction and learnability for outdoor augmented reality gaming, *in* 'Australasian User Interface Conference', Hobart, Australia, pp. 17–24.

Bauer, M., Heiber, T., Kortuem, G. and Segall, Z. (1998), A collaborative wearable system with remote sensing, *in* 'Int'l Symposium on Wearable Computers', Pittsburgh, Pennsylvania (find the abbreviation for this), pp. 10–17.

Biocca, F., Tang, A. and Owen, C. (2006), Attention funnel: omnidirectional 3d cursor for mobile augmented reality platforms, *in* 'Proceedings of the SIGCHI conference on Human Factors in computing systems', Montral, Qubec, Canada, pp. 1115–112.

Cheok, A. D., Goh, K. H., Liu, W., Farbiz, F., Fong, S. W., Teo, S. L., Li, Y. and Yang, X. (2004), 'Human pacman: a mobile, wide-area entertainment system based on physical, social and ubiquitous computing', *Personal and ubiquitous computing* **8**(2), 71–81.

Cutting, J. E. (2002), 'Reconceiving perceptual space', *Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space* .

Haller, M., Drab, S. and Hartmann, W. (2003), Real-time rendering: A real-time shadow approach for an augmented reality application using shadow volumes, *in* 'Symposium on Virtual reality software and technology', pp. 56–65.

Höllerer, T., Feiner, S., Terauchi, T., Rashid, G. and Hallaway, D. (1999), 'Exploring mars: developing indoor and outdoor user interfaces to a mobile augmented reality system', *Computer and Graphics* **23**(6), 779–785.

Jacobs, K., Nahmias, J.-D., Angus, C., Reche, A., Loscos, C. and Steed, A. (2005), Shadows: Automatic generation of consistent shadows for augmented reality, *in* 'Proceedings of Graphics Interface 2005', Victoria, British Columbia, pp. 113–120.

Kakuta, T., Oishi, T. and Ikeuchi, K. (2005), Shading and shadowing of architecture in mixed reality, *in* 'Int'l Symposium on Mixed and Augmented Reality', Vienna, Austria, pp. 200–201.

Krum, D. M., Omoteso, O., Ribarsky, W., Starner, T. and Hodges, L. F. (2002), Speech and gesture multimodal control of a whole earth 3d visualization environment, *in* 'Symposium on Data Visualisation', Barcelona, Spain, pp. 195–200.

Kurata, T., Oyabu, T., Sakata, N., Kourogi, M. and Kuzuoka, H. (2005), Tangible tabletop interface for an expert to collaborate with remote field workers, *in* '1st Int'l Conference on Collaboration Technology', Tokyo, Japan, pp. 58–63.

Kuzuoka, H. (1992), Spatial workspace collaboration: A sharedview video support system for remote collaboration capability, *in* 'Conference on Human Factors in Computing Systems', Monterey, CA, pp. 533–540.

Laurentini, A. (1994), 'The visual hull concept for silhouette-based image understanding', *Transactions on Pattern Analysis and Machine Intelligence* **16**(2), 150–162.

Li, M., Magnor, M. and Seidel, H.-P. (2003), Hardware-accelerated visual hull reconstruction and rendering, *in* 'Graphics Interface', Halifax, Nova Scotia, pp. 65–72.

Livingston, M. A., J. Edward Swan, I., Gabbard, J. L., Höllerer, T. H., Hix, D., Julier, S. J., Bailot, Y. and Brown, D. (2003), Resolving multiple occluded layers in augmented reality, *in* 'Int'l Symposium on Mixed and Augmented Reality', Tokyo, Japan, pp. 56–65.

Milgram, P. and Kishino, F. (1994), 'A taxonomy of mixed reality visual displays.', *IEICE Trans. Information Systems* **E77-D**(12), 1321–1329.

Piekarski, W. and Thomas, B. H. (2003*a*), Interactive augmented reality techniques for construction at a distance of 3d geometry, *in* 'Workshop on Virtual Environments', Zurich, Switzerland, pp. 19–28.

Piekarski, W. and Thomas, B. H. (2003*b*), An object-oriented software architecture for 3d mixed reality applications, *in* '2nd Int'l Symposium on Mixed and Augmented Reality', Tokyo, Japan, pp. 247–257.

Stafford, A., Piekarski, W. and Thomas, B. H. (2006), Implementation of god-like interaction techniques for supporting collaboration between outdoor ar and indoor tabletop users, *in* 'Int'l Symposium on Mixed and Augmented Reality', Santa Barbara, CA, pp. 165–172.

Sugano, N., Kato, H. and Tachibana, K. (2003), The effects of shadow representation of virtual objects in augmented reality, *in* 'Int'l Symposium on Mixed and Augmented Reality', pp. 7–10.

Thomas, B. H., Krul, N., Close, B. and Piekarski, W. (2002), Usability and playability issues for arquake, *in* 'International workshop on entertainment computing 2002', Makuhari, Japan, pp. 455–462.

Wither, J. and Höllerer, T. (2004), Evaluating techniques for interaction at a distance, *in* 'Int'l Symposium on Wearable Computers', Arlington, VA, pp. 124–127.

Wither, J. and Höllerer, T. (2005), Pictorial depth cues for outdoor augmented reality, *in* 'Int'l Symposium on Wearable Computers', Osaka, Japan, pp. 92–99.