

Implementation of God-like Interaction Techniques for Supporting Collaboration Between Outdoor AR and Indoor Tabletop Users

Aaron Stafford, Wayne Piekarski, and Bruce H. Thomas
Wearable Computer Laboratory
School of Computer and Information Science
University of South Australia
Mawson Lakes, SA, 5095, Australia
{aaron.stafford, wayne.piekarski, bruce.thomas}@unisa.edu.au

ABSTRACT

This paper presents a new interaction metaphor we have termed “god-like interaction”. This 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. Physical objects are captured by a series of cameras viewing a table surface indoors, the data is sent over a wireless network, and is then reconstructed at a real-world location for outdoor augmented reality users. Our novel god-like interaction metaphor allows users to communicate information using physical props as well as natural gestures. We have constructed a system that implements our god-like interaction metaphor as well as a series of novel applications to facilitate collaboration between indoor and outdoor users. We have extended a well-known video based rendering algorithm to make it suitable for use on outdoor wireless networks of limited bandwidth. This paper also describes the limitations and lessons learned during the design and construction of the hardware that supports this research.

CR Categories: I.3.6 [Computer Graphics]: Methodology and Techniques – Interaction Techniques; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism – Virtual Reality; J.9.e [Mobile Applications]: Wearable computers and body area networks.

Keywords: Outdoor Augmented Reality, Video-Based Rendering, Table-Top Interfaces, Indoor-Outdoor Collaboration.

1 INTRODUCTION

When performing specialised operations such as search and rescue, there are people located at indoor control rooms who typically oversee and manage the situation with a high-level understanding of the situation. People working out in the field act as sensors and have a more local understanding of the situation but may not be aware of the bigger picture as it develops. Currently, communication between participants is typically achieved using basic verbal communication via radio. With a large number of people in the field, managing this communication is quite complex, and tools to make this more effective are always desired. With the availability of outdoor augmented reality (AR) systems and indoor tabletop projector display systems, how can we combine these technologies to improve collaboration between indoor and outdoor users? AR systems are mainly focussed on presenting information to a single user, and to be used as a tool for collaboration, appropriate interaction metaphors and user interfaces need to be developed.

In this paper, we propose a new interaction metaphor *god-like interaction* that facilitates improved communication of situational and navigational information between indoor users equipped with tabletop displays, and outdoor users equipped with mobile AR systems. We have based our interactions on the capabilities of mythical god-like characters portrayed in popular culture such as

the computer game *Populous*, where a god-like creature controls a miniature isometric view of the world, and the mini-series *The Stand*, in which the hand of god comes out of the sky to interact with the environment. With our new AR metaphor, indoor users work with a tabletop display that is a miniature virtual representation of the outdoor world with the same top-down perspective a god-like character would possess. Indoor users use their god-like powers to point to objects in the tabletop world, and place physical props on the table to create new life-sized objects to be experienced by AR users outdoors. Outdoor users see these changes in real-time and experience these interactions as though they are



Figure 1 – An indoor user pointing at a location on the tabletop surface, which contains a virtual representation of the outdoor world. The table uses four video cameras to capture a 3D representation of objects located on the table.

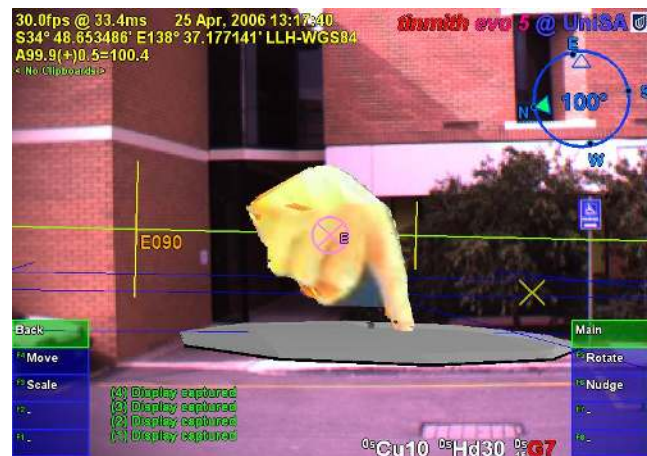


Figure 2 – Outdoor AR view showing the indoor user's hand appearing from the sky, and pointing toward a location in the distance. Voice commands from the indoor user tell the outdoor user specifically what task to perform.

appearing from the sky above. As depicted in Figure 1, an indoor user can provide navigational instructions by simply pointing to an area on the table surface and adding extra verbal commands. The result of this interaction for the outdoor users is a god-like hand that comes out of the sky to indicate a location as seen in Figure 2, and audio containing further instructions.

The indoor tabletop work area is comprised of a flat surface overlaid with projected information from a ceiling mounted projector, showing a virtual representation or satellite image of the outdoor environment. Surrounding the display area is an opaque wall with video cameras looking inwards, designed to capture objects and gestures that occur on the table surface. Indoor users can use their hands as well as physical props to describe situational information or navigational tasks. Objects and interactions captured by the cameras are scanned and sent wirelessly to outdoor users with mobile AR systems. For example, if an indoor user places a cup over the iconic representation of a building on the indoor table, a virtual giant cup will appear over the building in the physical world.

We have constructed a system that implements our novel god-like interaction metaphor, and this paper describes our implementation as well as a series of novel applications to facilitate collaboration between indoor and outdoor users. Rather than using traditional input devices, props, and trackers, our god-like interaction techniques use full 3D capture and reconstruction in order to preserve as much information about the indoor actions as possible. We use Li's Video-Based Rendering algorithm (VBR) [20] as the basis for the 3D reconstruction, which was designed to run on high-speed networks. Our outdoor users are mobile and connected via relatively slow wireless networks, and so we have extended Li's algorithm by pre-processing and compressing images on machines indoors to reduce bandwidth usage.

Furthermore, our techniques improve on existing work by combining a tabletop display with real-time 3D capture. Previous systems [24, 33] use large rooms with controlled lighting and wide-baseline camera set-ups [22] to perform 3D capture in an egocentric manner. Our system is constrained to a small table area where users interact with the system in an exocentric manner. This paper also describes the limitations and lessons learned during the design of the hardware to support this work, since there were a number of technical challenges to overcome to make this research feasible. With these combined novel contributions, this paper demonstrates the ability of indoor and outdoor users to supplement traditional voice communication with natural gestures and captured geometry.

2 RELATED WORK

There has been a wide range of work in the area of collaborative 3D environments such as AR and virtual reality (VR). Systems such as Studierstube [34] have been developed which allow multiple users to work together and edit 3D models in real-time, sharing a distributed scene graph between multiple application instances. With users present in the same room, they can use traditional forms of communication such as speech to coordinate their actions easily. When users are not present in the same location, collaborative tasks are more difficult and other methods of communication need to be explored.

Bauer *et al.* [2] developed a system that makes use of video and audio to provide a communication channel for people working remotely. With this system an expert watches a live video feed taken from a forward-looking camera mounted on the head of the remotely located worker. With both users seeing the same view of the world, the expert can guide the remote user through the task using simple audio commands.

There has been significant research into remote collaboration, specifically where an expert is working with remote field workers, such as Kurata *et al.* [15] and Höllerer *et al.* [10]. Much of this work focuses on users' interaction with the systems rather than interaction between the users. In a study performed by Kuzuoka [16], AR was used to convey gesture information for the purpose of remote instruction. The author found that gestures clearly increase communication efficiency and that gestures significantly reduced the number of verbal expressions required to convey intention.

Leibe *et al.* [18] created the *Perceptive Workbench* that uses vision based methods to identify and track objects on or above the table. As a result natural hand gestures can be used to interact with the system to perform such tasks as navigating around a terrain map. The authors also present a novel 3D reconstruction approach that makes use of multiple IR lights and only a single camera. The system is capable of constructing 3D geometry of objects placed on the surface, however the objects are not reconstructed in real time and because of the IR based reconstruction approach texture information is not available.

God-like interaction metaphors have been used previously to encapsulate the form of interaction between different types of users. In a system developed by Holm *et al.* [11], the authors use a Greek god metaphor to describe the relationship between a desktop user and a fully immersed VR user. The users work together to build a 3D environment. To the VR user, the desktop user appears as a giant hand interacting with the environment. The VR user is known as the hero and has the ability to "lift" massive objects that would not be possible in the real world. An interesting benefit of this work is that while both users design the environment, the immersed user also immediately experiences the environment and provides practical feedback about the usability to the desktop user. This system is limited to working with a set of prefabricated objects. This is similar to the work of Leigh and Johnson [19] who used the terms deity and mortal to describe the exocentric or egocentric roles assumed by the users of their system.

Nakanishi *et al.* [29] use the term transcendent communication to refer to the interaction between people with a bird's eye view of an area of interest and those at ground level in the area of interest. In a study conducted by the authors they found that the bird's eye view was effective for understanding the spatial movements of crowds. The authors also found that users with a bird's eye view were able to effectively assist people at ground level in understanding their surroundings. An interesting effect of working in god-like proportions was discovered by Zhang and Furnas [37]. In an informal user study the authors found that entities rendered at gigantic proportions were often intimidating to other users whose viewpoints were much closer to the ground.

In order to help to better understand interactions in different types of environments, Poupyrev *et al.* [32] created a novel classification for Virtual Environments (VE) manipulation metaphors. The classification separates metaphors into egocentric or exocentric depending on the user's viewpoint. Exocentric are those metaphors in which users have an external or god's eye view looking down onto the world. Egocentric metaphors are typically used in immersive systems and place the user directly in the environment. An example of a system which implements both of these metaphors is the Magic Book by Billingham *et al.* [3]. In this application, an AR user with a head mounted display holds a real book in their hands with a fiducial marker on the page. Using an exocentric view of the book, virtual scenes are overlaid onto the pages and the user can move the book to adjust the viewpoint. If a

reader finds a particular story interesting, they can assume an egocentric view of the world by replacing the real-world AR view with a virtual view and changing their scale to match the characters in the story. There has been a wide range of interaction metaphors designed to support users working beyond their reach, such as Scaled World Grab [27] and Worlds In Miniature (WIM) [36]. With these metaphors, a fully immersed user is able to change the size of the world to suit the task, and use it for navigation over long distances or manipulating large objects. Systems such as MultiGen SmartScene [28] support collaboration between multiple users using techniques such as this, with avatars of varying size indicating the user's location and scale to other users.

Grasset *et al.* [8] showed that exocentric-egocentric collaboration is significantly more efficient for navigational tasks compared to single person navigation. The user study made use of AR and VR for the exocentric view. While the authors recognise the potential benefit that a tangible interface could have for the exocentric user they did not do any research into this area. Interestingly they found that 60% of the users liked being able to see their hands while interacting in an egocentric manner.

Brown *et al.* [5] performed an evaluation of a shared mixed reality system. Users of the system shared the experience of walking through a museum together. The study observed the interactions of three users: a user in the museum with an exocentric view of the world on a PDA showing the position of the virtual collaborators, a fully immersed VR user with an egocentric view and an online user with a limited exocentric view of the museum. An interesting result of the evaluation found that users spent considerable time attempting to understand each other's limitations.

There has been a large amount of work into developing systems that can capture real scenes from multiple camera positions for the purpose of using the captured content in a different context [6, 9, 21, 23, 24, 33]. These systems are known as wide-baseline camera set-ups, because they have many cameras mounted on the outside of a scene pointing towards the centre. All of these systems are attempting to capture large scenes such as whole actors and the configuration generally is based around an entire modified room. Mobile implementations of these algorithms have also been implemented, such as a system developed by Ahrenberg *et al.* [1]. This system is based on a portable set of cameras that can be taken anywhere and set up in a small amount of time. The cameras look inwards and are able to capture events at the location where the event is taking place rather than holding the event in a specially designed studio. This system records synchronized video of the scene but processing is performed offline. Other systems such as those by Lok [21] are designed to improve performance by taking advantage of graphics hardware to perform the intensive computations. Another feature of Lok's work is that it captures the body of the user so that they can see their own body when fully immersed in VR. Lok supplemented a set of cameras with a head mounted camera to improve the rendering from the user's point of view.

Arbitrary physical props are commonly used to communicate information between two or more people [25, 26]. Physical props can be used to help express situations or scenarios by acting as placeholders for the real objects that they represent. The Virtual Round Table [4] extends this idea by overlaying the physical object with a virtual object that the physical object represents. The Virtual Round Table is a collaborative AR environment that is designed to support location-independent mixed reality applications. This system enables users to use physical objects such as cups as location markers for virtual objects. Physical props are visually identified using object recognition software. A head

mounted display can be used to overlay the virtual objects on top of the physical object. This system enables groups of people to annotate the physical environment by making use of arbitrary objects to represent significant items of interest.

There has been a wide range of research in the area of tabletop systems. An example is the tabletop system by Krum *et al.* [14] that accepts input from a number of different devices as well as hand gestures and speech. In an evaluation of input devices for the system [13] it was found that a mouse was preferred for navigating through the application. Speech was the next most effective device for navigation, and the least effective were gestures. The authors attribute the poor performance of the gestures to the recognition algorithms used, and not the actual use of gestures per se. A common interaction technique for tabletop displays is touch, such as the DiamondTouch [7] input technology that supports multiple inputs simultaneously. Touch-based technologies provide good affordance to co-located collaborators.

3 GOD-LIKE INTERACTION

The goal of our god-like interaction metaphor is to present users with a framework 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.

Our term god-like interaction embodies the way that god-like characters interact with our environment as perceived in popular culture. In computer games such as *Populous* from Bullfrog Productions and *Black and White* by Lionhead Studios users assume the role of a god-like character. They have a viewpoint in the sky and interact with the environment through an iconic hand controlled using a mouse. In *The Stand*, a mini-series based on a Stephen King book, there is a scene where the hand of a god comes out of the sky to detonate an explosive device.

The popular representation of god-like characters seems to have the common theme of a person floating above the world and looking down upon its occupants, comparable to a person looking down on a tabletop display. For this reason we use a tabletop to present the outdoor environment to indoor users. Tabletops have also been shown to be an ideal surface for group work as they support numerous people working equally together in a shoulder-to-shoulder arrangement which is not possible with wall-mounted displays [35]. Tabletops also support the ability to have physical props placed on them, and props have been shown to be useful because they are intuitive and efficient interfaces [12]. With the availability of powerful mobile AR systems, it is now possible for outdoor users to experience god-like interactions as though they are really happening in the physical world.

Based on the god-like framework for interacting with the system, this can be used to consider how people might perform actions specific to a search and rescue operation. An initial example is providing navigational information. Working within the framework an obvious approach would be to point at a location on the table and say "go here", as depicted in Figure 1. To tell someone outdoors to avoid a particular location an indoor user could simply circle the area with a hand gesture and say "don't go here". Another alternative is to write the word "Danger" on a Post-it note, stick it on a drink can and put it in the area of danger. To remove content from the environment, users can simply lift it off the table.

Outdoor users should have the ability to point at an object to select it, which would highlight the object on the indoor table. Users

could then ask questions to clarify something about the object in question. Outdoor users should also be able to translate, scale, rotate, or modify the geometry of objects to make changes that are reflected in real-time for indoor users. This would support the ability for indoor and outdoor users to collaboratively work together discussing the design of an object. Indoor users could provide the initial object shape from a prop, and the outdoor users can then modify it. Each user can perform tasks that are more suited to the location and user interfaces they have at their disposal.

It is important to realise that these are just example interactions that are immediately obvious. Any interactions that users are comfortable with that fit within the capabilities of the system are possible, and so the examples in this paper are not the only possibilities. These interactions project each of the users' actions and intentions. Any implementation must be kept as generic as possible to allow users to express themselves fully. The purpose of this framework is to provide a new set of capabilities that users can explore to find out what is more effective for their tasks.

4 PROTOTYPE APPLICATIONS

Based on our previous discussion, we believe that our god-like interaction metaphor is a compelling metaphor for remote communication since it is a concept that many people intuitively understand from popular culture. We have constructed a system to enable us to explore this interaction metaphor and to evaluate the effectiveness of the metaphor for remote communication. We informally trialled a number of different possibilities and describe some of the capabilities of our metaphor in this section.

4.1 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 as depicted in Figure 1 and speak a verbal command. The indoor user's hand is rendered at the corresponding physical location for people outdoors as shown in Figure 2. Indoor users can also point and drag their fingers to indicate a path to follow or a boundary to avoid crossing.

Simple props such as a jar from the kitchen cupboard can be used to highlight significant landmarks, or be used as virtual rally points or waypoint markers, as shown in Figure 3. The quality of the reconstruction is such that written text on the jar is legible and common objects are easily identifiable. The object in Figure 3 is clearly recognizable to most Australians as a jar of Vegemite. With easily recognizable objects, giving outdoor users directions is simplified with instructions such as "Go to the jar of Vegemite". Another example is shown in Figure 4, where a can of drink, which the indoor user happened to be using at the time, was transformed into a table prop simply by placing it onto the table. An advantage of the ability to use arbitrary props is that indoor users do not need to rely on specially prefabricated objects as props, allowing them to use the system with few restrictions. While almost all objects can be handled reasonably well, objects that have transparent areas will be difficult to subtract from the background, and if the object colour matches the background the system will not be able to differentiate it.

Rather than just using arbitrary objects as props, specific props can be used in some tasks to convey more information. Figure 5 shows how a miniature wooden street warning sign is used to warn outdoor users about wombats in the area. Although the object is quite thin, the multiple cameras are able to capture and reconstruct the object relatively accurately. If other objects such



Figure 3 – A jar of Vegemite used as an indoor table prop; the outdoor AR user sees it registered to the physical world.



Figure 4 – Using a commonly available indoor prop to place down a marker that is registered to the physical world.

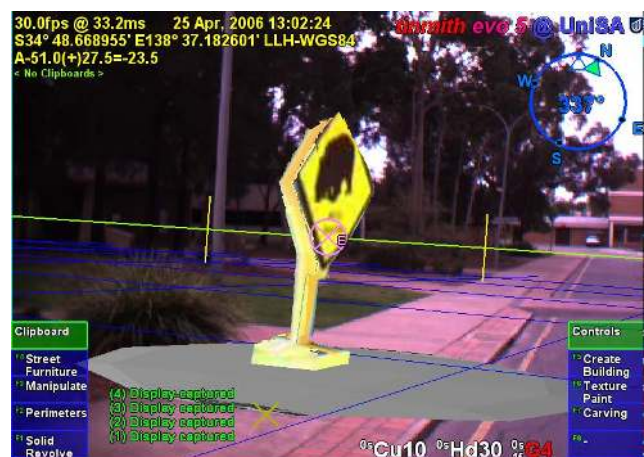


Figure 5 – An example of a toy street sign being used to warn outdoor users of numerous wombats in the area. as toy cars or animals are placed onto the table, they will appear as life-sized and realistic-looking models to outdoor users.

4.2 Augmented Post-It notes

During typical use there will be situations where it is handy to have persistent information accompanying a prop, which might vary from prop to prop as well. Indoors users can write information on Post-It notes and stick them to the surface of the props wherever it is convenient. The quality of the reconstruction pro-

vided is such that hand written notes are clearly legible to people outside. A block of Post-It notes can be kept close to the table, and indoor users can quickly and easily write and attach notes to objects on the table. 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.

In Figure 6 we show an example of how we have used our Augmented Post-It note technique to guide a user to find a buried treasure outdoors. The indoor user has taken the Vegemite jar from Figure 3 and added a Post-It note indicating what is buried underneath. The wombat sign from Figure 5 has been modified to include an arrow indicating exactly where the buried gold is located. The yellow Post-It note almost matches the colour of the sign, and so these notes appear to blend in quite well. With these Post-It notes, annotations can be made that are more permanent even when outdoor users are not in the nearby vicinity. Voice communications may require outdoor users to be able to see the objects being discussed, and over time voice communications will be forgotten if the task needs to be performed at a later time.

4.3 Relocation

Since the working volume on the table is limited, an external track ball mouse is used to scroll the map on the table. On the projected tabletop surface, a grey circular pad is drawn onto the map, indicating the region that scanned objects can be placed which will allow them to be scanned. The view of the tabletop system is such that the grey circular pad is always drawn at the centre of the display, and when the track ball mouse is used to scroll the map the grey circular pad remains in the centre of the table. Outdoor users also see a similar grey circular pad registered to the physical world, and as an indoor user scrolls the map around this grey pad will move around in the outdoor view as well since the systems are synchronised together. This relocation mechanism is analogous to flying across the landscape, and so fits into our god-like interaction metaphor. It also 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.

With the ability to do relocation, a number of interesting possibilities for interaction exist between the indoor and outdoor users. Indoor users are able to zoom the map in or out, which will affect the scale that objects are presented to outdoor AR users. An indoor user could place down a toy car with a "Follow Me" Post-It note, and then using the track ball mouse cause the car to drive across the landscape to guide outdoor users to the destination. Outdoor users also have the ability to interact with the reconstructed object and make changes. Using the modelling techniques built into Tinmith [31], outdoor users can select the object using special gloves and then manipulate the object. If an outdoor user does not like the position that the object is placed at, they can simply move it to the desired location. Figure 7 shows an example of where a user selects an object and then lifts it up into the sky. The movement operation is using our AR working planes techniques in head-relative mode, so that the object maintains its distance from the user while being manipulated. These changes are reflected in real-time back to the indoor users, and cause the moving map to adjust so that the grey circular pad is in the centre of the table. Indoor users are then able to make further adjustments to the object, or even move it along to give the impression that the

cat is flying across the sky. Indoor users currently see grey pads representing reconstructed objects since there is little or no information about the reconstructed objects when looking at it them from above. Future work will look at alternate avatars that can be used instead of the grey pads that will convey more information about the objects that they represent.

4.4 Snapshot

With a limited working volume, placing more than one object on the table at any time is not practical unless the objects are quite small. This is a severe limitation because complex tasks will typically require more than one object to appear at any time. Also, the reconstruction algorithm is best used for capturing single objects with convex hulls, and not the capture of multiple arbitrary objects scattered across the table because objects can easily block the visibility of others from the cameras. At the moment there is no solution to these capturing problems so we added features to the user interface to make our desired complex interactions possible. To allow multiple objects to be visible at once, we added a simple snapshot feature that enables copies to be made of the currently reconstructed object. 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. The live reconstruction is still occurring at the same location, and using the previously described relocation technique the live version can be moved away from the snapshot. Now two objects are both visible at the same time, one of them is a snapshot from the past, and the other is a live version from the table. Figure 6 was captured using the snapshot technique.

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. Figure 8 shows a scene where a number of objects have



Figure 6 – Existing props from previous examples augmented with Post-It notes to convey extra meaningful information about the environment.



Figure 7 – The outdoor user is able to manipulate reconstructed objects and move them to different locations (1) Outdoor user about to select the cat using the gloves to begin manipulation of the cat, (2) Outdoor user lifting the cat into the sky using an AR working plane in head coordinates.



Figure 8 – External orbital view of the outdoor AR user, showing a number of snapshot objects appearing on the virtual landscape, along with an avatar representing the user.

been placed over the landscape. The grey circular pad was moved into the desired location, and the snapshot function activated. The grey circular pad is then moved to the next location, and the snapshot activated again. The same reconstruction can be copied multiple times, or new props can be used as desired. An advantage of the snapshot tool is that the same prop can be reproduced many times without having to have multiple physical instances of the prop available. We envisage a number of interesting applications for the snapshot function. One application could be using a single prop to instantiate multiple copies of a breadcrumb trail on the ground for people to follow to get to a target location. Another example we envisage is a collaborative town planning application, where miniature toy houses, cars, street signs, and traffic lights could be placed down and manipulated by either indoor or outdoor users. Since the objects are represented with realistic looking props, the level of abstraction is much closer to what is experienced in the real world, and both indoor and outdoor users will have no trouble understanding their meaning.

5 OUTDOOR AR AND INDOOR TABLE-TOP SYSTEM DESIGN

We use our Tinmith mobile outdoor AR system [30] 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. The tabletop system is comprised of two separate systems: the object capture software broadcasts the information necessary to reconstruct objects; and the table display software renders a top-down representation of the outdoor world.

It is necessary to separate users' gestures from the rest of their bodies, and so we have built a tabletop device resembling a miniature movie set, as can be seen in Figure 1. The table has a blue-coloured wall perimeter, so that props placed onto the table are easily segmented from the background. We use four Point Grey Dragonfly cameras evenly spaced and all pointing towards the centre of the table. The cameras are mounted using standard optical camera mounts to support fine-grained calibration. 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. The cameras are kept horizontal to simplify the calibration process.

Reconstruction using VBR is only possible on the volume above the table that is visible by all cameras, and we term this space the working volume (v_{wa}). The cameras' field of view as well as the radius of the table dictates the size of the working volume. With an infinite number of cameras around the table, the working volume is approximated as a cylinder with a cone capping the top, as depicted in Figure 9. With fewer cameras the working volume will become increasingly polygonal with a slightly larger working volume. The radius of the base of the working volume r_v is trigonometrically derived from half of the horizontal field of view of the cameras θ_h . The heights of the cylinder h_{cy} and cone h_{co} are trigonometrically derived from half of the vertical field of view of the cameras θ_v . The final working volume v_{wa} is the sum of the cylinder and cone volumes and is calculated using: $v_{wa} = \pi r_v^2 (h_{cy} + \frac{1}{2}h_{co})$

The viewpoint of the cameras mounted on our table are 50 cm from the centre, that is $r_i = 50$. The Dragonfly cameras with a 4 mm lens have an effective horizontal field of view of 70 degrees, giving $\theta_h = 35^\circ$, and an effective vertical field of view of 50 degrees, giving $\theta_v = 25^\circ$. Using the equation specified above, the working volume of our table is $= 42945 \text{ cm}^3$. This volume is much smaller than the overall capacity of the table, and therefore it is important to plan the design of the table carefully to maximise the working volume.

There are limitations in how the table can be expanded to support a larger working volume. Increasing the field of view or moving the cameras further away reduces the resolution of the capture and requires the wall height to be increased to block out the outside world. If a small wall height is desired, the wall may be so close to the centre that the cameras cannot fit inside. We use a sloped wall so that it is at the desired height but with the base widened to include the cameras.

6 RECONSTRUCTION ALGORITHM

VBR algorithms such as those by Li *et al.* [20] and Matusik *et al.* [23] perform many of the reconstruction steps in a similar way. A series of cameras placed around the object are used to capture images. The background is removed from these images, and then a vector-based contour is calculated that expresses the outline of the object. Since the camera locations are known, these contours can be projected from the camera locations into space. These projected contours intersect at the location where the physical object is, and using constructive solid geometry (CSG) techniques the volume that is shared by all the projected contours can be calculated and then rendered. This volume is known as the visual hull [17] and is an approximation of the physical object.

The calculations required to perform VBR can be very CPU intensive if performed in a naïve fashion, and so novel algorithms have been developed which attempt to accelerate this process. The algorithm by Matusik *et al.* [23] takes advantage of certain properties of the contours to calculate a polygonal mesh with textures of the final object in less time than the naïve method. This method is currently the best implementation when a true 3D mesh is required for further processing, but is highly intensive for a CPU to process at real-time rates. However, algorithms by Lok [21] and Li *et al.* [20] both take advantage of the processing power of modern 3D graphics cards, and perform the visual hull reconstruction within the GPU at a significantly faster rate. Both algorithms make use of projective textures to render the visual hull and do not actually compute a polygonal mesh. Lok makes use of the stencil buffer and requires a pass for each input image, while Li *et al.* is more efficient and uses one texture unit per camera image and alpha blending in a single pass.

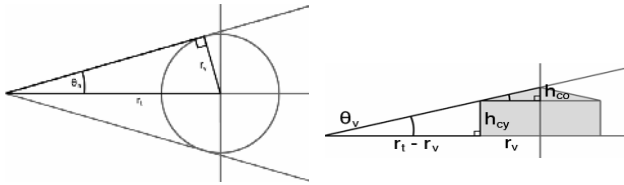


Figure 9 – Top-down and side-on views of the working volume made up of cylinder and cone portions controlled by the field of view of the cameras.

All of the outlined methods require the camera images to be sent over the network, to be used as textures so that the reconstruction appears as realistic as possible. Therefore, the main difference between the approaches is that Matusik *et al.* would require the transmission of the final 3D mesh from an indoor server while Li *et al.* requires the transmission of the image contours as a series of vectors along with a compressed alpha mask for each image. The transmission of the contours and alpha mask consumes much less bandwidth than even an optimised 3D mesh.

The Nvidia GeForce Go 6600 used in Tinmith has four texture units restricting reconstruction to four cameras, but as newer graphics hardware with more texture units become available this will permit more cameras. We use four easily transportable Mac Mini computers, each dedicated to processing the image from a single camera. Each Mac Mini processes captured images to remove the background, and produce a smoothed alpha mask with small holes removed. The OpenCV library is used to calculate a vector-based contour that approximates the outline of the foreground object.

While previous implementations of VBR use powerful machines on local networks, we are required to use wireless networks with orders of magnitude less bandwidth. We use the FFmpeg library to compress the RGB images for transmission by a factor of 40, and separately compress the alpha mask using run-length encoding. The contour data is very simple and is sent over the network without any compression. We transmit updates as UDP packets over the network for each camera at five frames per second. Since the VBR reconstruction is performed locally on the mobile computer, the outdoor person can continue to view the visual hull from any viewpoint even when the network fails.

The rendering loop within the Tinmith software running on the mobile system is synchronised to the camera that captures the live AR view of the world, which operates at 30 frames per second. When the reconstruction algorithm is running for a single object, there is no noticeable speed decrease and the algorithm runs within the frame rate of the system. As objects are copied using the snapshot feature, the frame rate gradually decreases, although the actual rate achieved depends on the number of objects that are actually visible at the time, not the total number that are being rendered. Figure 8 demonstrates how with nine snapshot objects in the scene graph, three of them just outside the field of view on the right, the frame rate has dropped to 14.7 frames per second. We have run the system with up to 20 objects at interactive frame rates, but this varies depending on the visibility of the objects and the size they appear as on the image plane.

7 FUTURE WORK

In this paper, we have presented an overview of our god-like interaction metaphor as well as some examples in use, and we are interested in trialling the system with real-world problems. We plan on evaluating the god-like interaction metaphor with organisations that would gain the most from these capabilities, and make improvements based on their comments. We also plan on per-

forming formal user studies to compare various approaches and existing techniques to decide which are the most effective.

There are a number of ideas that we plan to explore using the table once we have improved the size of the working volume. One example that we think would be interesting would be to drive a small remote-control car around the table, causing the car to appear to drive around over the landscape outdoors. Another example would be to have robotic props move themselves around the table according to changes made by outdoor users. We believe that our god-like interaction metaphor opens up a wide range of possibilities that are still waiting to be explored.

8 CONCLUSION

This paper has presented our new interaction metaphor that we have termed *god-like interaction*. We have developed this metaphor in order to improve collaboration between users located outdoors using mobile AR systems, and users located indoors working on tabletop projected displays. The metaphor is based around the ability of indoor users to place physical props as well as portions of their bodies onto a table surface that is continuously scanned in real-time as a 3D object and then transmitted over a wireless network to remote users.

In this paper we demonstrated a number of example techniques that we believe take advantage of our god-like interaction techniques. Using hand gestures and commonly available physical props, our system supports the conveyance of navigation instructions to outdoor users. Using our augmented Post-It notes, extra information may be added to objects for reference at a later time. The relocation technique helps indoor and outdoor users to work around the limited working volume of the table and control the placement of the reconstructed object. The snapshot feature allows multiple copies of objects to be made so that only one prop needs to be on the table at any time, and the reconstruction algorithm is able to scale to a large number of objects.

We explained the design process behind the construction of our table and techniques that can be used to expand the working volume. We have extended an existing VBR algorithm by Li *et al.* [20] by adding various pre-processing and compression stages to allow it to work over a low-bandwidth and unreliable wireless network. Our implementation is able to run in real-time on a mobile outdoor AR system, and has allowed us to perform informal evaluations of our techniques. We believe that our system is useful in supporting rich collaboration between indoor and outdoor users, and that since the god-like interaction framework is very generic, it will allow users to explore new interactions that have not been previously thought of by the designers.

9 ACKNOWLEDGEMENTS

The authors would like to thank the Apple University Consortium for supplying our Mac Mini computers, the Australian Defence Simulation Office for sponsoring the development of our mobile backpack systems, and the Division of ITEE and the University of South Australia for their financial support.

10 REFERENCES

- [1] Ahrenberg, L., Ihrke, I. and Magnor, M. A Mobile System For Multi-Video Recording. In *1st European Conference on Visual Media Production*, pp 127-132, London, UK, Mar 2004
- [2] Bauer, M., Heiber, T., Kortuem, G. and Segall, Z. A Collaborative Wearable System with Remote Sensing. In *2nd Int'l Symposium on Wearable Computers*, pp 10-17, Oct 1998
- [3] Billingham, M., Kato, H. and Poupyrev, I. *The MagicBook: Moving Seamlessly between Reality and Virtuality*. Computer Graphics and Applications, Vol. 21, No. 3, pp 6-8, 2001.

- [4] Broll, W., Meier, E. and Schardt, T. The Virtual Round Table - a Collaborative Augmented Multi-User Environment. In *3rd Int'l Conference on Collaborative Virtual Environments*, pp 39-45, San Francisco, CA, Sep 2000
- [5] Brown, B., MacColl, I., Chalmers, M., Galani, A., Randell, C. and Steed, A. Lessons From The Lighthouse: Collaboration In A Shared Mixed Reality System. In *Human Factors in Computing Systems*, pp 577-584, Ft. Lauderdale, FL, Apr 2003
- [6] Debevec, P., Yu, Y. and Borshukov, G. Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping. In *9th Eurographics Rendering Workshop*, Vienna, Austria, Jun 1998
- [7] Dietz, P. and Leigh, D. DiamondTouch: a multi-user touch technology. In *14th Symposium on User Interface Software and Technology*, pp 219-226, Orlando, FL, Nov 2001
- [8] Grasset, R., Lamb, P. and Billinghurst, M. Evaluation of Mixed-Space Collaboration. In *4th Int'l Symposium on Mixed and Augmented Reality*, pp 90-99, Vienna, Austria, Oct 2005
- [9] Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., Koller-Meier, E., Svoboda, T., Gool, L. V., Lang, S., Strehlke, K., Moere, A. V. and Staadt, O. *blue-c: a spatially immersive display and 3D video portal for telepresence*. Transactions on Graphics, Vol. 22, No. 3, pp 819-827, 2003.
- [10] Höllerer, T., Feiner, S., Terauchi, T., Rashid, G. and Hallaway, D. *Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system*. Computer & Graphics, Vol. 23, No. 6, pp 779-785, 1999.
- [11] Holm, R., Stauder, E., Wagner, R., Priglinger, M. and Volkert, J. A combined immersive and desktop authoring tool for virtual environments. In *Virtual Reality*, pp 93-100, Mar 2002
- [12] Ishii, H. Tangible Bits: Coupling Physicality and Virtuality Through Tangible User Interfaces. *Mixed Reality: Merging Real and Virtual Worlds*. Y. Ohta and T. Hideyuki. Tokyo, Japan, Ohmsha Ltd, pp 229-247, 1999
- [13] Krum, D. M., Omoteso, O., Ribarsky, W., Stamer, T. and Hodges, L. F. Evaluation of a Multimodal Interface for 3D Terrain Visualization. In *Conference on Visualization*, Boston, MA, Oct 2002
- [14] Krum, D. M., Omoteso, O., Ribarsky, W., Stamer, T. and Hodges, L. F. Speech and Gesture Multimodal Control of a Whole Earth 3D Visualization Environment. In *Symposium on Data Visualisation*, pp 195-200, Barcelona, Spain, May 2002
- [15] Kurata, T., Oyabu, T., Sakata, N., Kourogi, M. and Kuzuoka, H. Tangible Tabletop Interface for an Expert to Collaborate with Remote Field Workers. In *1st Int'l Conference on Collaboration Technology*, pp 58-63, Tokyo, Japan, Jul 2005
- [16] Kuzuoka, H. Spatial workspace collaboration: A sharedview video support system for remote collaboration capability. In *Conference on Human Factors in Computing Systems*, pp 533-540, Monterey, CA, May 1992
- [17] Laurentini, A. *The Visual Hull Concept for Silhouette-Based Image Understanding*. Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 2, pp 150-162, 1994.
- [18] Leibe, B., Stamer, T., Ribarsky, W., Wartell, Z., Krum, D., Weeks, J., Singletary, B. and Hodges, L. *Towards Spontaneous Interaction with the Perceptive Workbench, a Semi-Immersive Virtual Environment*. Computer Graphics and Applications, Vol. 20, No. 6, pp 54-65, 2000.
- [19] Leigh, J. and Johnson, A. E. *Supporting Transcontinental Collaborative Work In Persistent Virtual Environments*. Computer Graphics and Applications, Vol. 16, No. 4, pp 47-51, 1996.
- [20] Li, M., Magnor, M. and Seidel, H.-P. Hardware-Accelerated Visual Hull Reconstruction and Rendering. In *Graphics Interface*, pp 65-72, Halifax, Nova Scotia, Jun 2003
- [21] Lok, B. Online Model Reconstruction for Interactive Virtual Environments. In *Symposium on Interactive 3-D Graphics*, pp 69-72, Chapel Hill, NC, Mar 2001
- [22] Magnor, M. A. *Video-Based Rendering*, A K Peters, 2005
- [23] Matusik, W., Buehler, C. and McMillan, L. Polyhedral Visual Hulls for Real-Time Rendering. In *12th Eurographics Workshop on Rendering Techniques*, pp 115-126, Jun 2001
- [24] Matusik, W., Buehler, C., Raskar, R., Gortler, S. J. and McMillan, L. Image-Based Visual Hulls. In *27th Int'l Conference on Computer Graphics and Interactive Techniques*, pp 369-374, New Orleans, LA, Jul 2000
- [25] McGee, D. R., Cohen, P. R. and Wu, L. Something from nothing: augmenting a paper-based work practice via multimodal interaction. In *Designing Augmented Reality Environments*, pp 71-80, Elsinore, Denmark, Apr 2000
- [26] McGee, D. R., Pavel, M., Adami, A., Wang, G. and Cohen, P. R. A visual modality for the augmentation of paper. In *Workshop on Perceptive User Interfaces*, pp 1-7, Orlando, FL, Nov 2001
- [27] Mine, M. R., Fredrick P. Brooks, J. and Sequin, C. H. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *24th Conference on Computer Graphics and Interactive Techniques*, pp 19-26, Los Angeles, CA, Aug 1997
- [28] Multigen. *SmartScene*. <http://www.multigen.com>
- [29] Nakanishi, H., Koizumi, S., Ishida, T. and Ito, H. Transcendent Communication: Location-Based Guidance for LargeScale Public Spaces. In *Conference on Human Factors in Computing Systems*, pp 655-662, Vienna, Austria, Apr 2004
- [30] Piekarski, W. and Thomas, B. H. An Object-Oriented Software Architecture for 3D Mixed Reality Applications. In *2nd Int'l Symposium on Mixed and Augmented Reality*, pp 247-257, Tokyo, Japan, Oct 2003
- [31] Piekarski, W. and Thomas, B. H. Interactive Augmented Reality Techniques for Construction at a Distance of 3D Geometry. In *Workshop on Virtual Environments*, pp 19-28, Zurich, Switzerland, May 2003
- [32] Poupyrev, I., Weghorst, S., Billinghurst, M. and Ichikawa, T. *Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques*. Computer Graphics Forum, Vol. 17, No. 3, pp 41-52, 1998.
- [33] Prince, S., Cheok, A. D., Fabriz, F., Williamson, T., Johnson, N., Billinghurst, M. and Kato, H. 3D Live: Real Time Captured Content for Mixed Reality. In *Int'l Symposium on Mixed and Augmented Reality*, pp 7-13, Darmstadt, Germany, Oct 2002
- [34] Schmalstieg, D. and Hesina, G. Distributed Applications for Collaborative Augmented Reality. In *Virtual Reality*, pp 59-66, Orlando, FL, Mar 2002
- [35] Ståhl, O., Wallberg, A., Söderberg, J., Humble, J., Fahlén, L. E., Bullock, A. and Lundberg, J. Information exploration using The Pond. In *4th Int'l Conference on Collaborative Virtual Environments*, pp 72-79, Bonn, Germany, Sep 2002
- [36] Stoakley, R., Conway, M. J. and Pausch, R. Virtual reality on a WIM: interactive worlds in miniature. In *SIGCHI Conference on Human Factors in Computing Systems*, pp 265-272, Denver, CO, May 1995
- [37] Zhang, X. and Furnas, G. W. Social Interactions in Multiscale CVEs. In *4th Int'l Conference on Collaborative Virtual Environments*, pp 31-38, Bonn, Germany, Sep 2002