A Hand-Held AR Magic Lens with User-Perspective Rendering

Domagoj Baričević*   Cha Lee†   Matthew Turk‡   Tobias Höllerer§   Doug A. Bowman¶

Four Eyes Lab   Center for Human-Computer Interaction
University of California, Santa Barbara   Virginia Tech

ABSTRACT

In this paper we present a user study evaluating the benefits of geometrically correct user-perspective rendering using an Augmented Reality (AR) magic lens. In simulation we compared a user-perspective magic lens against the common device-perspective magic lens on both phone-sized and tablet-sized displays. Our results indicate that a tablet-sized display allows for significantly faster performance of a selection task and that a user-perspective lens has benefits over a device-perspective lens for a selection task. Based on these promising results, we created a proof-of-concept prototype, engineered with current off-the-shelf devices and software. To our knowledge, this is the first geometrically correct user-perspective magic lens.

Keywords: User-perspective view, magic lens, user study, prototype, MR simulation, augmented reality

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology; I.3.0 [Computer Graphics]: General

1 INTRODUCTION

A magic lens is generally considered to be a see-through interface/metaphor that affords the user a modified view of the scene behind the lens. This concept was first conceived for traditional desktop applications [4] but has been applied to Virtual Reality (VR) [30] and Mixed Reality (MR) systems. In a sense, all Augmented Reality (AR) displays are magic lenses, in that the virtual content can only be viewed through the display. In the early years of AR research and prototyping, both head-worn and hand-held solutions were demonstrated, perhaps with a slight predominance of head-worn displays in the 1990s. In recent years, hand-held AR demonstrations and applications have become commonplace as smart devices have permeated the market. Smart phones and small tablets now come equipped with tracking sensors, cameras, powerful GPUs, and high resolution displays. The integration of all these technologies has allowed developers and researchers to create real AR applications on small, phone-sized and tablet-sized devices for the masses. *For the rest of this paper, we will refer to hand-held AR displays as AR magic lenses.*

Recent technological advances bring about the opportunity to correct a limitation with current AR magic lenses: the device-specific (as opposed to user-specific) perspective of the augmented scene. Many existing concept images of AR magic lenses show that the magic lens displays a scene from the user's perspective, as if the

---
*e-mail: domagoj@cs.ucsb.edu

†e-mail: chalee21@cs.ucsb.edu

‡e-mail: mturk@cs.ucsb.edu

§e-mail: holl@cs.ucsb.edu
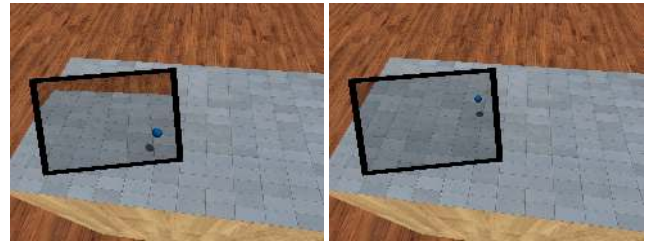
¶e-mail: dbowman@vt.edu

Figure 1: A side-by-side comparison of a device-perspective magic lens (left) and a user-perspective magic lens (right) as simulated in our VR system. The image on the device-perspective lens is rendered from the viewpoint of the camera situated on the back of the display. The image on the user-perspective lens is rendered from the viewpoint of the user. *This is best viewed in color.*

display were a smart transparent frame allowing for perspective-correct overlays. This is arguably the most intuitive view. However, the actual magic lens shows the augmented scene from the point of view of the camera on the hand-held device. The perspective of that camera can be very different from the perspective of the user, so what the user sees does not align with the real world. Figure 1 compares user-perspective and device-perspective magic lenses. We define the user-perspective view as the *geometrically correct view of a scene from the point-of-view of the user, in the direction of the user's view, and with the exact view frustum the user should have in that direction*. There has been previous work in creating an augmented view based on the user's perspective [11], but the final imagery does not achieve the full user-perspective effect. We know of no existing hand-held AR display that properly renders a geometrically correct user-perspective image. This is due to the difficulties which must be overcome to generate user-perspective imagery.

There are three challenges in presenting user-perspective imagery: (1) track the user's head position accurately, (2) track the pose of the display with respect to the world accurately, and (3) obtain/create an accurate model of the scene. Generating user-perspective images requires all three components, since we must render the world from an arbitrary observer viewpoint different from the device's viewpoint. Until recently, it has not been possible to tackle all three problems with hand-held devices, but newly emerging technologies bring the promise of success. Many smart devices now come equipped with front-facing cameras which can be used to track the user's head pose with respect to the device [1]. These devices also provide back-facing cameras and sensors which can be used to track the device with respect to the world via computer vision techniques [31, 13, 20]. What is probably the most challenging issue is building a model of the world in unprepared and dynamic environments. Although techniques have started to appear [23, 19, 12], it is still challenging to perform robust, accurate 3D reconstruction on mobile devices. However, we expect these capabilities to be widely available in the near future.

The goal of our work is to determine if a user-perspective AR magic lens is useful, beneficial, or desirable for users. We seek to address these questions: Does a user-perspective magic lens provide significant benefits over the traditional device-perspective

magic lens in AR applications? Does such a device provide intangible benefits for the user experience? To answer these questions we need to evaluate and compare a user-perspective magic lens against its traditional device-perspective counterpart. We claim that it is not possible to adequately evaluate user-perspective magic lenses with currently available technology. The results from such a study would be affected by confounding factors, mostly caused by inadequate 3D reconstruction techniques. These factors include jitter, visual artifacts, frame rate and robustness. Instead, we conduct the study using MR Simulation [16].

With MR Simulation, a high-fidelity VR system is used to simulate an MR display with controllable levels of fidelity. Although the use of simulation also has limitations, we argue it is a valid approach for this experiment. We seek to compare two different types of views rather than two existing magic lenses. We would like to give each view type equal treatment, and a real prototype would not do justice to the user-perspective magic lens. While we cannot yet build a real user-perspective magic lens with sufficient image quality and performance, we can use MR Simulation to deliver the correct view of such a display for evaluation.

In this paper we report our findings on a user study using MR simulation to compare four different magic lenses, varying device size (phone-sized and tablet-sized) and perspective (user-perspective and device-perspective). We show that for a selection task in a small AR workspace, the user-perspective imagery and tablet-sized displays provide significant benefits over the device-perspective imagery and the phone-sized displays. Encouraged by the results of our study we implemented a proof-of-concept user-perspective magic. We describe our prototype, created with off-the-shelf hardware (Kinect and Wiimote) and open source software. We discuss the limitations of the device, given our implementation choices, and also future possibilities for such displays.

2 RELATED WORK

The magic lens interface is a well known visualization metaphor dating back to the early 1990s on desktop display systems. It was first initially conceived by Bier et al. at Xerox Parc [4] as a 2D filter which would reveal hidden information, enhance data of interest, or suppress distracting information. This concept allowed users to move a virtual window on the display which would reveal or suppress information directly beneath the virtual window. This was a powerful tool that was soon extended to 3D display systems. Flat lenses and volumetric lenses were used in immersive VR systems (such as presented by Viega et al. [30]) enabling better visualizations that could allow users to view hidden data inside of 3D objects as well as the virtual world surrounding them. Schmalstieg et al. [27] used magic lenses (via transparent props) for a workbench VR system, allowing users to create a rich pallet of tools on the magic lens and affording an intuitive two-handed interface. Handheld AR systems can easily be thought of as magic lenses [5, 18] and are probably the most prolific AR system with the advent of smart phone devices [25].

The argument could be made that any AR display is a "magic lens" as AR displays allow users to view virtual content otherwise hidden in the real world. Optical see-through HMDs used in AR act as a view encompassing magic lens [15]. Although optical see-through HMDs present a perfect user-perspective view of the real world, the virtual content may still be incorrect and contain some parallax error. The AR community has previously looked at parallax free HMDs [29, 28, 7]. These HMDs are configured such that objects on the display surface properly align with the objects in the real world backdrop. Takagi et al. [29] contributed a stereo, optical HMD with zero parallax and State et al. [28] built a parallax-free video-see-through HMD.

Mobile AR is still a relatively new area of computing and presents many unique UI challenges. For the most part the community has been focused on solving the technical challenges of AR, since the technical obstacles are still large and have not changed significantly since they were first recognized [2, 3]. Researchers have envisioned AR as a form of ubiquitous computing [9, 6], supporting many users in large indoor and outdoor environments. Until recently the proposed systems were of the wearable type such as backpack AR systems. With the rise of powerful smart devices and available communication networks in most urban areas, handheld or pocket AR systems are now possible. Most commercial AR applications are either an AR browser (Layar, Wikitude, etc.) or based on image recognition (Google Goggles, StikyBits, ShopSavvy). A recent survey by Olsson et al. [22] measured the overall acceptance and user experience with these types of applications. Their results indicated an inconsistent user experience and raised concerns regarding practical usefulness and technical unreliability. This is one indication that we do not fully understand how to design AR applications and that the technology is still very much in its infancy. We need to better understand how to leverage current and near-future technologies to create useful and enjoyable mobile AR experiences. Our goal in this paper is to determine whether user-perspective magic lenses are beneficial.

We would like to know what can affect user performance and to gather information on the overall user experience with magic lenses. Researchers have looked into these questions via simulation [21, 17] and via real AR magic lenses [25, 26]. Oh et al. [21] looked at the effects of magic lens display size and screen aspect ratio on a search and a path following task. Their results indicated that, in general, a larger magic lens was better for searching while a smaller magic lens was better for path following. Their findings made sense since the authors employed user-perspective magic lenses. In this scenario, a larger lens would result in a larger field of view (FOV) and more available content to search at any instant, allowing for faster search times. On the other hand it could also occlude more real-world context, thus slowing down path following. In Looser et al. [17], the authors compared different selection techniques in a virtual workbench environment. Direct-touch, ray-casting, and a virtual pointer (the magic lens) were tested on a selection task and the magic lens was found to be more enjoyable and faster than either direct-touch or ray-casting. Rohs et al. [25, 26] have done much work on using real mobile phones as AR display systems. Their results indicate that target acquisition using magic lenses follow a modified Fitts' Law using both laboratory data (2D maps) and real world data (3D outdoor scenes).

The various possible causes of perceptual problems with magic lens interfaces are well known in the AR community. A recent summary of these perceptual issues can be found in Kruijff et al. [14]. Hill et al. [11] investigated the issue of motion parallax with the concept of "virtual transparency". Here they used fisheye cameras to capture both a wide FOV view of the scene in front of the display and a view of the user. With face tracking they determined the view frustum the user should have of the AR scene. This information was then used to transform the fisheye view from the back-facing camera so that it approximates the view the user should be seeing. This approximate frustum still has the camera's point of view and is only aligned with the correct frustum at a specified target distance; objects at other distances appear either smaller or larger than they should and incorrect occlusion between objects at different distances still occurs. Some objects may disappear while others may be duplicated: objects closer than the target distance may fall outside the view even though they should be visible by the user, objects further than the target distance may appear on the display even though they should be outside the view. These frustum misalignment issues are most obvious in near-field AR where the distance between the user and the display is similar to the distance between the camera and the workspace. However, with far-field AR the distance between the user and display can be negligible com-

Table 1: A comparison of a user-perspective magic lens in simulation against a real world version with current technology, with respect to their fidelity in terms of several immersion factors (elements of the perceived experience). In general, the simulated system affords visuals with less artifacts, more robust tracking, and unrestricted interaction. The real world version affords more realistic physical and perceptual stimuli but is limited by visual artifacts and tracking performance.

Immersion Factor	Fidelity of Simulation	Fidelity of Real AR
User-perspective view	Geometrically correct, no visual artifacts	Geometrically correct, with **visual artifacts**
Interaction	Affords unrestricted interaction	**Range, speed limited by technology**
Tracking	Robust, jitter free, no mis-registration	**Less robust, has jitter, possible image mis-registration**
Display resolution	Lower, dependent on simulator	Higher, dependent on magic lens
Image realism[1]	Low resolution, low geometric detail, inaccurate lighting and shadows	Realistic, with **visual artifacts**
Environment realism[2]	Low resolution, low geometric detail, inaccurate lighting and shadows	Full
Visual perception	Restricted FOV, inaccurate stereo overlap, vergence-accomodation mismatch	Full
Proprioception/kinesthesia	Some latency, possible mis-registration of relative position of magic lens to user	Full
Physical fidelity	Incorrect weight, dimensions, tactile feedback	Full

[1] The realism of the scene seen through the magic lens
[2] The realism of the world outside the magic lens view

pared to the distance from the camera to the scene background. If this approach is applied to panoramic AR, the frustum misalignment should be relatively small and the end result could be a good approximation of a user-perspective magic lens.

To render a user-perspective image on the magic lens, the application requires a model of the scene and accurate tracking within that scene. With the recent advances in sensor technology and computer vision techniques we may soon achieve this. Pollefeys et al. [23] demonstrated a complete system for building visual models using hand-held cameras. Klein et al. [13] presented a real-time system for tracking a hand-held camera in small AR workspaces. The PTAM (parallel tracking and mapping) system is able to track a camera in unprepared environments with no a priori information while also able to produce a 3D map of the environment. More recently Newcombe et al. [19, 20] and Izadi et al. [12] published methods for reconstructing dense models using a single camera and the Kinect sensor, respectively. These methods allow us to create detailed models with per-pixel 3D information of arbitrary environments. Although the methods require bulky sensors and more processing power than most hand-held devices currently have, we can reasonably expect to see smaller sensors and more powerful devices in the near future.

Until the technologies required for a user-perspective magic lens have matured enough for practical applications, we need a different approach if we want to evaluate these magic lenses. One approach is to use Mixed Reality (MR) simulation [10, 24, 16, 8]. MR Simulation is a method which utilizes a VR display system to simulate a variety of MR displays for the purpose of conducting user experiments. It has been used to evaluate text legibility with AR displays in outdoor environments [10], registration error and system latency on AR systems [24, 16], and tracking recovery methods for hand-held AR systems [8]. With MR Simulation we can simulate the displays we desire and evaluate them in a controlled manner. Our paper takes this approach and the next section reports on our study comparing user-perspective and device-perspective magic lenses.

3 COMPARING A USER-PERSPECTIVE MAGIC LENS AND A DEVICE-PERSPECTIVE MAGIC LENS

One possible approach to building a user-perspective magic lens would be to use a transparent screen for optical-see-through AR. Using a transparent screen would introduce other significant issues such as binocular rivalry and visual interference as users try to align virtual objects rendered on a transparent layer against objects in the background [15]. We experimented with a mock-up of a user-perspective magic lens, consisting of a piece of clear glass with stickers for annotations, and found it difficult to align the stickers with objects in the background.

In our simulator, we take a computer vision approach which requires a 3D model of the environment to correctly render a scene from any viewpoint different from the device. Unfortunately it is not yet possible to create a user-perspective lens with the same visual fidelity as a device-perspective lens using this approach. Although the technology has advanced enough that we can develop a proof-of-concept prototype such as the one discussed later in this paper, we would not be able to create a system with the level of performance required for the kind of evaluation we are looking to do. The robustness of the tracking and modeling and the quality and detail of the model are too poor for the system to be used as our evaluation testbed.

From an evaluation standpoint we would like to evaluate a "perfect" user-perspective view on a magic lens and compare it against a perfect device-perspective view on the same display. The best solution available at this point is to simulate our lenses. Using MR Simulation offers more experimental control and replicability and is a practical way to evaluate non-existing display systems. With the whole experiment defined in software we can quickly try out different variations of our experimental task, including changing anything in the scenery with just a small change in the code. This also lets us easily adapt our environment to the different users if necessary, e.g., lowering/raising objects depending on the user's height. Table 1 illustrates the differences of a simulated and real AR agic lens.

In the case of this evaluation, simulation also affords us the ability to create a user-perspective lens with the same visual fidelity as a device-perspective lens, thus allowing for a fair comparison. Moreover, MR Simulation gives us a simulated AR system that can perform better than any current real-world system. The registration of the virtual content is pixel-perfect with absolutely no latency or any kind of error, which even the best computer-vision methods cannot yet achieve. The difficult problem of 3D reconstruction is non-existent in simulation as all the geometry in the scene is known and completely under our control.

Of course, the use of MR Simulation does come with some limitations, inherent in any VR system. Table 1 gives an overview of some of the perceptual and technical factors involved, and how they compare between simulation and a real AR system. These factors have to be taken into account when designing any VR experiment to ensure that they do not unduly affect the results. In this instance, the most important aspect is to ensure that the correct view is given by both lens types, while removing visual artifacts. Our interest is only in the effect of perspective on user performance with a magic lens, something we assume to be a fundamental property of human spatial reasoning and kinesthesia. Thus, we expect that the trends observed with a unbiased simulation with carefully controlled visual properties to also be true for the real-world case. Until this is possible to confirm with a real-world evaluation, simulation will remain the best way to gain insight into this problem.

In our preliminary analysis, we looked at what advantages a user-perspective lens may have over a device-perspective lens and came to the conclusion that a user-perspective lens would perform better when spatial manipulation is involved, for example in cases where users have to reach into view of the lens to select or interact with an object. During this action a user would need to reach into the world behind the lens while viewing this action through the lens itself. Similar to how refraction in water affects how we estimate the position of objects under water, we hypothesized that a device-perspective lens would also reduce our ability to accurately perform this action. The act of reaching into the world comprises three parts. First the user must search for a target. Then the user must reach in and reach for the target before finally manipulating it in some manner. The most interesting part of this action is the second part, as this is where the user's understanding of where the virtual target exists with respect to their hand is crucial. We predicted that it would be during this short time that the user-perspective magic lens would outperform a traditional, device-perspective lens. We therefore designed an experiment with a direct selection task, which emphasizes this particular action.

In a pilot study we refined our experimental design and the task we would ask users to perform. We simulated four different hand-held magic lenses by changing two different parameters: display type (user-perspective and device-perspective) and display size (phone-sized and tablet-sized). We decided on a task which would require the user to select nearby objects in rapid succession while forcing the user to move their hand into and out of the view of the lens after each selection. This would emphasize the selection part of the task and provide more data on selection performance. We chose to also manipulate the size of the display to determine whether the effects of perspective rendering would change with respect to display size, since the perceived FOV of the perspective lens would also change as the display size changed. We created our hypotheses for how the users would perform using each lens and how they would rate their preference for each lens:

H1: Users will be better at selecting virtual objects with a user-perspective magic lens than a device-perspective magic lens.

H2: Users will be better at selecting virtual objects with a larger magic lens than a smaller magic lens.

H3: Users will prefer a user-perspective magic lens over a device-perspective magic lens.

3.1 Simulation, Environment, and Task

Simulating the Magic Lenses   To conduct this experiment, we simulated a small workspace and two types of magic lenses in four different configurations, using MR Simulation and an immersive VR display system. We simulated two user-perspective and two device-perspective magic lenses; one phone sized and one tablet sized for each type. We simulated these magic lenses as simple flat rectanglular shapes with a textured quad as the display surface (as seen in Figure 1). The imagery of the two display types were generated differently.

The simulated device-perspective lenses assumed that the images shown on the display were rendered from the point of view of the camera at the back of the simulated device. The camera had a 45° VFOV and was centered horizontally and near the top of the back of the device. To create the image displayed on the simulated device-perspective lenses, we first rendered the scene from the camera's viewpoint to an off-screen buffer. We then applied this image as a texture to the lens' display surface.

To create the user-perspective magic lenses, we instead rendered the scene from the head-position of the user (centered between the user's eyes) onto an off-screen buffer. Then using a projective texturing approach, we applied the correct region onto the display surface of the lens to create the user-perspective image. It is important to note that the image on the simulated user-perspective lens is a monocular image taken from the single point between the user's eyes and will not be perfectly aligned with the background.

The differences between both lens types can be seen in Figure 1. To use the magic lenses, the user wore a head-tracked HMD and held a 6DOF tracking device in one hand. This provided the pose of the magic lens and the user's head position. During each frame the correct lens image was generated, applied to the simulated magic lens, and then the user was presented with a stereo view of the scene. It is important to note that while the simulation was in stereo, the image on the magic lens was always in mono as this is what would happen with a real magic lens display of these types.

We did not simulate the user's hands in this experiment, since we did not think we could simulate the hands with enough fidelity. Our VR system and tracking system could not robustly track and render a realistic hand for the user. We decided that not having hands would be better than having visually unrealistic hands that could disturb the users. To allow the users to directly select the virtual targets, we placed a tracked pointer in the user's other hand. Users used this device to perform the selection and did not use their fingers or hand directly.

Environment   We created a simple virtual room with a plain bench as the environment for this user study. As seen in Figure 2a, the user was placed next to a table that had a large ball by the side. The ball was placed on side of the table that was on the user's dominant hand. The top of the table was adjusted to approximately the height of the user's stomach as to make reaching the targets easier. The top of the table was textured with a simple gray-cement-like texture to make searching more difficult. Next to the table was a large ball, initially white, which the user needed to interact with to proceed through the various phases of the task (this ball would also change color to indicate the current phase). We used a model of a virtual art gallery as the backdrop for the room. As seen in Figure 2b, the simulated magic lens (a traditional device-perspective lens in this case) was primarily held in the user's off hand although users were given the option of choosing either hand. The small blue ball in the figure was the 3DOF pointer which was held in the user's dominant hand and was used to manipulate the objects in the scene.

A few important features need to be discussed. Since the simulated magic lenses are intended to represent AR magic lenses, we designed them to also have the same artifacts as video-see-through AR displays. One key requirement is that all virtual (augmented) content appear on top of the "real" content. As seen in Figure 2c, the virtual orange target appears in front of the blue pointer (a real object). This effect would also occur in real world AR applications and can result in users wrongly believing a virtual object is in front of a real object. We chose to keep this feature for both the user-perspective and device-perspective lenses. Next is how the

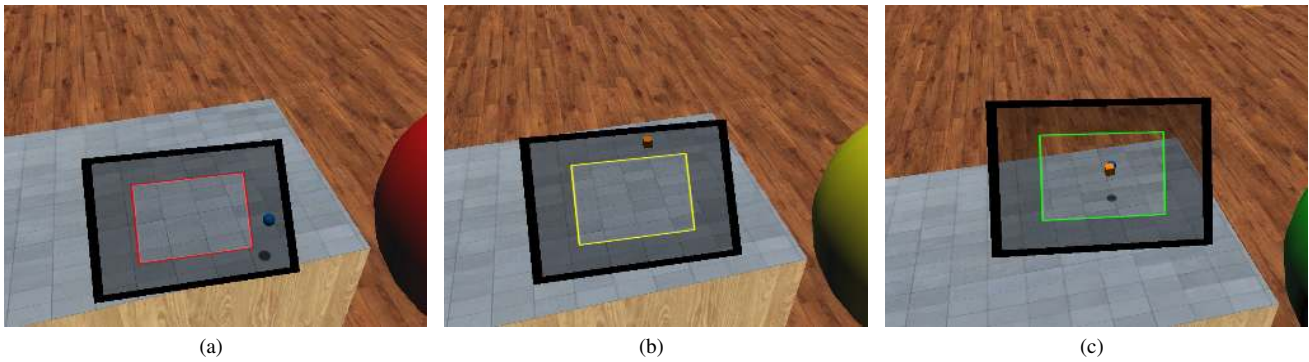|		
(a)	(b)	(c)

Figure 2: Screenshots of various phases during the task that the user was asked to perform using the user-perspective magic lens. These images are in mono while the experiment was conducted in stereo. (a) shows the wait phase, indicated to the user by the red color of the big ball and outline of the sub-window on the magic lens. The target is not visible in this phase. (b) shows the search phase, indicated by the yellow color of the big ball and outline of the sub-window. The target is now visible if seen through the magic lens. (c) shows the selection phase, indicated by the green color of the big ball and outline of the sub-window. Notice the occlusion of all simulated "real" objects by the virtual target and the shadow cast by the blue pointer.

shadows were rendered. No virtual objects cast shadows while all real objects do cast shadows. As seen in Figure 2c, the blue pointer has a shadow while the virtual target does not.

While these two requirements make the selection task more difficult, they are reflective of what happens in most AR applications that do not have a model of the environment. This is true for the device-perspective lens. Although the user-perspective lens assumes a model of the world which could be used to create correct occlusion and shadows, we chose not to do this to make the experiment fair for both lens types. The last point is that while the environment was rendered in stereo, the content which appeared on the lens was always rendered in mono. For the device-perspective lens this was rendered from the point of view of the device. With the perspective lens this was rendered from the user's head position (between their eyes).

Task To keep users interested we created a game for this study. The user was asked to perform a simple selection task. Similar to Whack-A-Mole, users were asked to repeatedly touch a virtual target. This single virtual target (represented by an orange box in Figure 2b and 2c) would appear randomly somewhere on the surface of the table and could only be seen via the magic lens. The goal of the game was to find and touch as many targets as possible within two minutes. To touch the target, users would move the blue pointer until it intersected with the orange target.

During the game, the user had to repeatedly search for a virtual target and select the target when it was found. We were mainly interested in how well users could select the targets, so we designed the game to decouple the two different tasks. We designed the game so the act of finding and selecting a target consisted of three different phases: wait, search, and selection. The user had to interact with the large ball by the table to progress from one phase to the next, with the color of the ball indicating the current phase.

The game's default state was the wait phase (shown in Figure 2a). During this phase the virtual target was not yet visible. This was indicated to the user by the red color of the big ball and red color of the outline of the sub-window on the magic lens.

To begin the search phase, the user had to place the blue pointer on or inside the big ball and maintain contact. This would activate the search phase (Figure 2b) and was indicated by the yellow color of the big ball and the yellow color of the outline of the sub-window on the magic lens. The virtual target was now visible if it was seen through the magic lens and the user would need to view the target through the sub-window of the magic lens to move to activate the selection phase.

Once the target was found and viewed through the sub-window, the big ball and the outline of the sub-window would turn green indicating to the user that the selection phase had begun (Figure 2c). Now the user could remove the blue pointer from the big ball and use it to select the target. If the blue pointer was removed from the big ball before the target was found, the game would revert back to the wait phase. Once the user had successfully selected the target, the game would revert to the wait phase and the user could attempt the next target.

3.2 Apparatus

Our VR display system consisted of an NVis SX111 wide-FOV HMD. This HMD provides a combined $102°$ horizontal FOV with 1280x1024 resolutions for each eye. Tracking of the users head and the simulated magic lens was provided by InterSenses IS900 PCTracker. The InterSense wireless wand was held in the user's off hand and provided the pose for the magic lenses. Since we only had two sensors for this device, we also included a WorldViz PPT tracking system to track the pointer device. A single marker was attached to a wand to provide the 3DOF position of the blue pointer. The VR software package used was the WorldViz Vizard development kit and ran on a Windows 7, 32 bit machine with 2GB of RAM and a 2.4 GHz Intel processor. The video-card was an NVIDIA Quadro FX 5600. This setup was sufficient for rendering the scene at 60 FPS for both eyes with stereo.

3.3 Experimental Design

We used a mixed design for our study, with two factors: display size and display type. Display size was the between-subjects factor and consisted of a phone-sized and tablet-sized design. The dimensions of the phone-sized lens was 7.5 by 5 cm while the dimensions of the tablet-sized lens was 22.5 by 15 cm. Display type was the within-subjects factor and consisted of a user-perspective magic lens and a device-perspective magic lens. With the user-perspective lens all content on the display was rendered from the user head position. With the device-perspective lens the content was rendered from the camera position on the back of the display. Every user was asked to use both display types. Each user was put through a training session followed by two timed sessions for each display type. The training session was at most 8 minutes long with randomly generated targets. The two timed sessions consisted of three games each. Considering both display types, each user was asked to play 12 games. We generated a set of 12 games where each game consisted of 100 randomly placed targets on the table. These same 12 games were

played by every user. Ordering for both display type and game were fully balanced with Latin squares and each game lasted 2 minutes long.

The dependent variables for this study were selection time and selection path deviation. We were not interested in the search time and did not explicitly measure that. Selection time was the measure of how long (in seconds) each person took to select a target, once it had been found, on each display type. To measure selection time, we started a timer as soon as the virtual target had been found and the selection phase began. The target was considered found when it could be seen in the sub-window of the magic lens. We stopped the timer as soon as the target was selected. The reported selection time was then the average of all selections times for every target the user successfully selected for each display type (over six games).

Selection path deviation is harder to define. We wanted a measure of how accurately (spatially) the users could move the pointer to select the target. If users were asked to touch an object in the real world, then their hand would travel a certain distance during this action. If we then asked the user to do this same action while viewing it through a display, we might expect a different total distance and that difference in the total distances would be a measure of how the display affected the selection task. The data from our pilot study showed that users generally moved their had in a small arc from the big ball to the target during the selection phase. However we did not observe a pattern we could use to determine the "ground truth" distance for each user based on this data, since it was different for each user. We instead chose to use the straight-line distance from the initial position of the pointer to the target as the optimal distance. Based on the results of our pilot study we also knew the deviation from the straight-line path to be small (less than a meter in most cases) since the object was at reaching distance. We normalized the deviation by taking the total distance the pointer traveled during the selection phase and dividing it by the straight-line distance from the original position of the pointer to the target. We called this the selection path deviation and it is a measure of the magnitude of the error in the path traveled by the selection hand (the blue pointer).

3.4 Participants and Procedure

Participants   We had 52 paid participants from a university subject pool. Participants were paid a nominal fee of $10 for their time. Of the 52 participants, 48 were included in the final data. Data from four users were removed due to system errors or VR sickness. All participants were tested for common color blindness via color vision plates obtained on-line and for stereo vision via random dot stereograms. Any participant who did not pass the tests was not included in the collected data. All participants also reported they had correct or corrected-to-normal vision.

Of the 48 users included, 24 were male and 24 were female, and they were evenly balanced between all conditions. We were surprised by the number of users who had very little experience with a smart device. 14 of the 48 users reported they had never owned a smart phone before, while 39 of the 48 users had never owned a tablet device. Considering that our users came from a university subject pool this was larger than we expected. Our users also had little experience with 3D graphics and VR/AR interfaces with most users (34 of 48) reporting they had never played 3D video games or only tried it briefly beforehand which was again surprising considering our participants came from a college student body. Most users also reported minimal or no prior experience with VR and AR applications. In summary, our users could be described as young (18-57, average of 20.6 years) with minimal experience with VR/AR and minimal experience with hand-held devices.

Procedure   At the start of the study each participant was given a questionnaire to obtain some background information regarding their experience level with the technologies involved in this study.

Each user was then given a simple color blindness test via color vision plates we obtained on-line. Next, the study administrator would take the user into the tracking space. The user was fitted with the HMD and given a simple stereo vision test via random dot stereograms. Any users which did not pass both the color blindness test and stereo vision test were disqualified and were thanked and paid for their time. The HMD was taken off and the study administrator would then give a brief explanation of the type of magic lens the user was about to see, how the game worked, and how to use the devices. After this explanation the training phase would start. During this first training phase, the study administrator would talk the user through each phase of the game. Once both the user and the study administrator were confident that the user understood the game, the user was told to practice until they were comfortable with the task. During this time the study administrator would observe and correct the user if necessary. Some users would randomly wave the wand around to select the target, which was a strategy that we did not allow. Once the training session ended the user was given a break, where the HMD was taken off and the user could sit down and relax. After this break the first timed session began. The user was told to attempt to select as many targets as they could within the 2 minute time limit while trying to be as accurate as possible. After each game in the session (3 games per session) the user was told to take a break if needed. If the study administrator observed any signs of VR sickness, a break was enforced and after each session a break was mandatory. Users were told they could quit at any time if they felt dizzy or motion sick. Once both timed sessions for the first display type were complete, the user took a longer break before repeating the same steps for the second display type. At the end of the trials, the user was given a post-questionnaire to obtain qualitative feedback and then was paid, thanked, and excused.

3.5 Results

Selection Time   The results for selection time can be seen in Table 2 and Figure 3.5. The large user-perspective magic lens performed the best overall and the small user-perspective performed the worst overall. Using an ANOVA, we found a large (ges = 0.3070) significant effect of display size on selection time (F(1,46) = 24.4318, p < 0.0001). In general users were able to select the targets much faster with the larger display than the smaller display with an overall average of 1.61 secs for the large display compared with 2.22 secs for the small display. The data did not show a significant effect from display type on selection time but the users with the user-perspective display did perform marginally faster with an overall mean of 1.88 secs over the users with the device-perspective display (1.95 secs). There was also a small (ges = 0.0196) interaction between display size and display type (F(1,46) = 5.5324, p < 0.0231). The user-perspective magic lens was better than the device-perspective lens when the display size was large, but worse when the display size was small.

Selection Path Deviation   When looking at selection path deviation, an ANOVA found a small (ges = 0.0039) significant effect of display type (F(1,46) = 5.8285, p = 0.0198). The user-perspective magic lens was better with an average deviation of 41.3% while the device-perspective magic lens had an average of 45.7% deviation in the distance traveled by the users selection pointer. We did not find a significant effect of display size on average path deviation nor an interaction between display type and display size on path deviation, although on average the users with the larger display performed better with a deviation of 39.0% against the users with the smaller display at 48.0%.

Post-Questionnaire   We asked users to complete a post-experiment questionnaire to gauge how they felt during and after the experiment. One of our concerns was with respect to their comfort level during the study. We asked users if they felt any dizziness or motion-sickness during the trials and most users reported

Table 2: Statistical results from an ANOVA using a mixed design of display size and display type on selection time. Generalized Eta Squared (η_G^2) is used as the effect size.

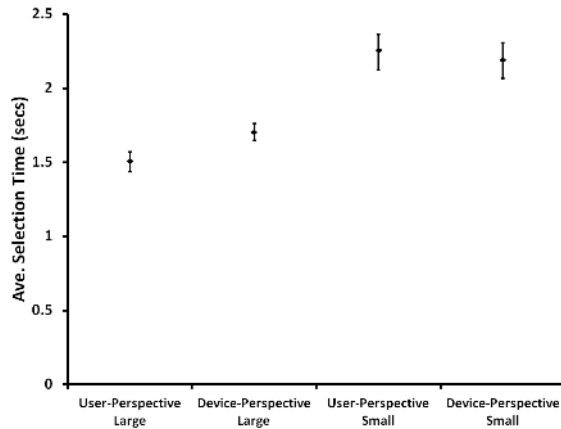| Effect	df	F	p	η_G^2
Display Size	1/46	24.4318	< 0.0001	0.3070
Display Type	1/46	1.6541	0.2048	0.0059
Size x Type	1/46	5.5324	0.0231	0.0196

Figure 3: Plot of means and standard errors for the average selection time. Selection time is expressed in seconds and is measured from the start of the selection phase until a target is selected and averaged over all targets for each condition.

Table 3: Statistical results from an ANOVA using a mixed design of display size and display type on selection path deviation. Generalized Eta Squared (η_G^2) is used as the effect size.

| Effect	df	F	p	η_G^2
Display Size	1/46	0.7634	0.3868	0.0158
Display Type	1/46	5.8285	0.0198	0.0039
Size x Type	1/46	0.0187	0.8918	0.0001

Figure 4: Plot of means and standard errors for selection path deviation. Path deviation is expressed as the total distance travelled by the pointer over the straight line distance for each target. The final percentage reported is the average of all targets.

little or no dizziness (46 out of 48). Only two users reported moderate dizziness during the trials. However some of the users did report physical discomfort. Approximately half of the users reported some discomfort with their eyes (23/48), neck (29/48), or back (19/48). We had expected this to be an issue as the HMD was relatively heavy (1.3 kg), and we designed the experiment to reduce the amount of time each user had to spend with the HMD on. Each user had been told to take as many breaks as they felt necessary and we enforced multiple breaks during the trials. Overall we did not notice any large issues with ergonomics and comfort levels during the trials, although numerous users did comment on the HMD as being uncomfortable in their written feedback.

We also asked the users to give feedback on the task itself. Most users did not find the task difficult with 31 finding it very easy while 16 found the task moderately easy and 1 found it difficult. We asked users to report their preferred device on three different aspects: overall, for finding the targets, and for selecting the targets. The results can be seen in Figure 5.

When asked which type of magic lens they preferred overall, all 24 users with the small magic lenses preferred the device-perspective display (Figure 5a). The majority (15/24) of the users of the large magic lenses preferred the user-perspective display while 7 preferred the device-perspective and 2 did not have a preference. This trend was also reflected in the results when we asked which display type was preferred for searching for the targets. All 24 users of the small magic lens preferred the device-perspective display while the users with the large magic lens were more even. 9 users preferred the user-perspective, 8 the device-perspective, and 7 had no preference (Figure 5b).

Both of these results are understandable. When the display size is small, a user-perspective lens generally has a much smaller FOV than the equivalent device-perspective lens. The larger FOV enabled users with the device-perspective lens to view a larger area
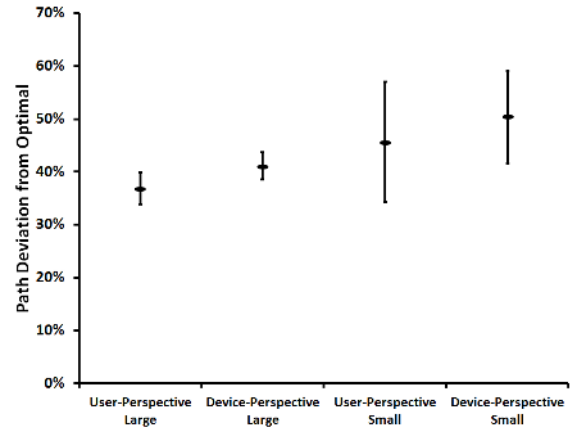
of the table and resulted in significantly faster search times. This directly affected this group's overall preference as the users spent more time searching than selecting, and the faster search times resulted in better scores. At the end of each 2 minute game, users were shown the score and the users averaged 26.4 targets using the user-perspective lens and 32.2 targets using the device-perspective lens. With this being the only feedback it is natural that users would prefer the device which they scored higher on.

When asked which magic lens they preferred for selecting the targets, the users generally preferred the user-perspective lens over the device-perspective lens (Figure 5c). The users with the small lens were more balanced in their preferences with 10 for the user-perspective, 12 for the camera-perspective, and 2 with no preference. The users with the large magic lens favored the user-perspective display. 13 users preferred the user-perspective, 5 preferred the device-perspective, and 6 had no preference.

3.6 Discussion

In our original hypotheses, we first predicted (H1) that users would be better at selecting the targets with a user-perspective magic lens than with a device-perspective magic lens. This hypothesis was based on the assumption that a user-perspective view would afford better spatial understanding for the user as it is the more natural view and closer to what a viewer would naturally see if his or her view was not mediated by any display. The results from our study provide some evidence to support this hypothesis. Although display type did not have a significant effect on selection time, it did have a small significant effect on path deviation.

This small effect can be explained by considering the task and how the display types affect the users' performance. In general the only time that selection performance is affected by display type is during a very short transition period when the pointer crosses from outside the lens' view and appears on the display. When the pointer

Overall Magic Lens Type Preferences

(a)

Magic Lens Type Preferences for Searching for Targets

(b)

Magic Lens Type Preferences for Selecting Targets
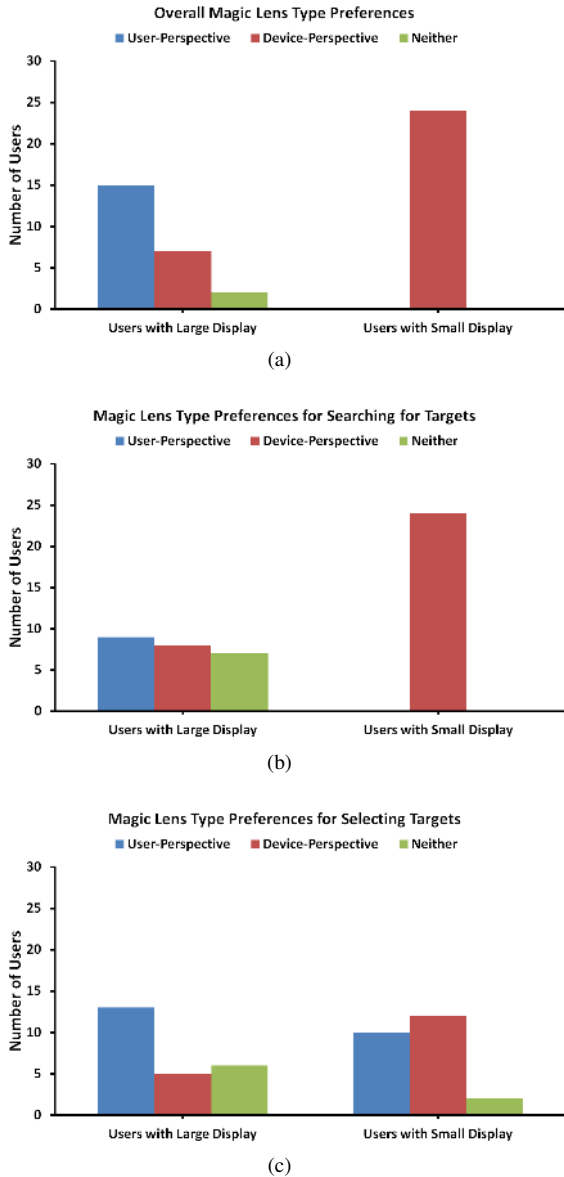

(c)

Figure 5: Results from post-questionnaire pertaining to display preference. (a) shows the results for overall display preference. (b) shows the results for display preference for searching for the targets. (c) shows the results for display preference for selecting the targets.

is outside the view of the lens or once the pointer is inside the lens' view for a significant time, the user can quickly adjust their spatial understanding of where the pointer is simply by the feedback of seeing the location of the pointer. What remains then is that short transition period when the pointer crosses from outside the view of the lens and into the view of the lens. Here the user may need to adjust the path the pointer is traveling as the pointer may not be exactly where the user expects it to be. This may not even be a large correction or require any real recognition on the user's part. This small effect that we observed is then understandable given the constraints of the task.

Our second hypothesis (H2) was that a larger display would also improve selection performance. While display size did not have a significant effect on selection path deviation it did show a large significant effect on selection time. This can be clearly seen in Fig-

ure 3.5. Users were able to select objects much faster using the larger tablet magic lens. Even though there was no significant effect on path deviation, the users with the larger display also averaged slightly better on that metric.

Our last hypothesis (H3) was that users would prefer the user-perspective magic lens over the device-perspective lens. In our post-questionnaire, we asked users to choose which lens they preferred overall for searching for targets and for selecting targets. While users overall preferred the device-perspective lens over the user-perspective lens, this result stems overwhelmingly from the users that operated with the small magic lens (Figure 5a). This was most likely due to the advantages of a higher FOV, which enabled faster search times and a higher game score, which also led to the small display users choosing the device-perspective lens for searching (Figure 5b). However these users also noticed that the perspective lens was better for selecting the objects, as seen in Figure 5c. This supports our hypothesis since we are mainly concerned with the selection phase. Although the users' responses were clearly affected by the goals of the game, this response is a reflection on how users may respond in real world AR applications. In small work environments, the need to search and look around for virtual objects is much reduced as users can quickly build a mental map of the environment. We believe that in these instances a user-perspective lens would be preferred since it also enables better selection performance. In outdoor environments it may be better to use a device-perspective lens as it affords a larger view of the world with most camera lenses and enables faster searching. Applications like Layar or Wikitude might actually suffer from using a user-perspective lens as it would restrict the amount of content which can be viewed due to the smaller FOV from the users' viewpoint.

4 A USER-PERSPECTIVE MAGIC LENS PROTOTYPE

The results of our study demonstrated the possible advantages an ideal user-perspective lens has over a device-perspective lens. We next wanted to determine how well current state-of-the-art technology could be used to implement such a display. We knew an optical-see-through approach would not work due to issues with aligning monoscopic augmentations with stereoscopic backgrounds; we had determined this by experimenting with both the physical mock-up mentioned above and with a virtual system in our simulator. Our prototype would therefore have to rely on re-rendering reconstructed 3D models of the real-world scene. We had considered using stereo reconstruction and/or structure from motion methods as they did not require special hardware. However, we felt that they were not yet fast and accurate enough for our needs. Instead we chose to go with a Kinect sensor as it offered accurate and robust reconstruction, at interactive rates.

4.1 Implementation

Our implementation is based on the KinectFusion algorithm, which provides high quality 3D reconstruction and also offers pose tracking. The pose tracking allowed us to establish a real-world coordinate system which we use to freely add augmentations anywhere in the space with no need for markers. We used an open source implementation of the KinectFusion algorithm (KinFu), provided by developers of the Point Cloud Library. We made some modifications to KinFu, specifically to the ray-caster which renders the reconstructed model, adding support for rendering the reconstructed volume both in color and from arbitrary viewpoints.

Fast and accurate user tracking was achieved through the use of infrared LEDs. This type of tracking would only work in environments with low ambient IR illumination, but the Kinect already imposed that restriction. We have the user wear goggles with four IR LEDs, which are then tracked with a Wiimote mounted on the top of the prototype and facing the user (see Figure 6a).

<table>
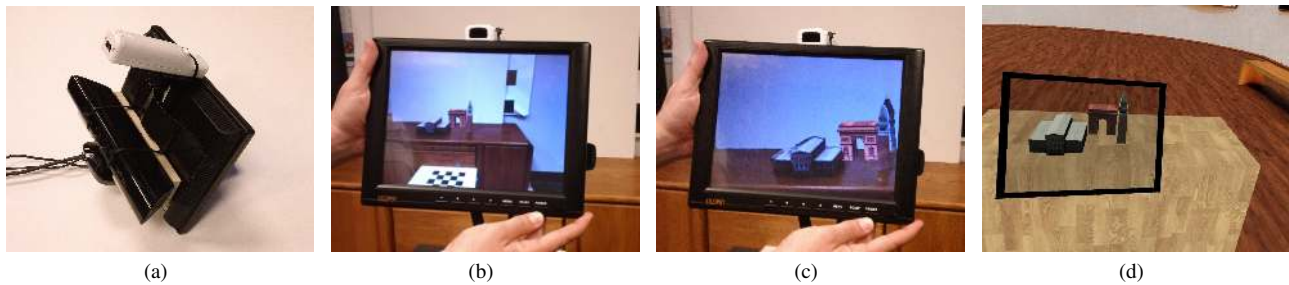<tr><td>(a)</td><td>(b)</td><td>(c)</td><td>(d)</td></tr>
</table>

Figure 6: A proof-of-concept prototype of an AR magic lens using a Kinect sensor. (a) shows the hardware used: a Kinect sensor, Wiimote, and tablet display. The Kinect sensor is mounted behind the tablet facing out and the Wiimote is mounted above the display facing the user. (b) shows a device-perspective magic lens. The image on the magic lens is from the perspective of the Kinect sensor. (c) shows the user-perspective magic lens. The image on the magic lens is from the perspective of the user. (d) shows what an ideal user-perspective magic lens would look like in our simulated environment.

The user study indicated that tablet-sized magic lenses are where user-perspective views have the greater advantage so we use a display of that size, a Lilliput 10.4" TFT LCD screen with a native resolution of 800x600 pixels. Our prototype is a tethered hand-held system, with the Kinect and the display connected to a workstation that does all the computation. The workstation runs Kubuntu and has 32GB of RAM, two dual-core AMD Opteron 2.60GHz CPUs, and two NVIDIA GPUs: Quadro FX 5600 and Quadro 6000.

4.2 Prototype Discussion

The type of result we can get with our prototype can be seen in Figure 6c. To render the AR scene, we use three graphical layers that are blended together. The base layer is the colored rendering from the modified KinFu raycaster, this layer has very limited detail due to the relative low resolution of the reconstructed colored volume but it can fill out areas that are not currently in view of the Kinect or where the live data has gaps. The mid layer is a reprojection of the point cloud from the live depth feed, this layer has much higher detail as it uses the full live color data from the Kinect sensor but it can have gaps and it is constrained to what the Kinect sees, not what the user sees. The final layer has the augmentations to the scene, rendered with proper occlusion with the real world. However, the occlusion handling is based on the KinFu model and the coarseness of the KinFu reconstruction can result in small occlusion errors, particularly along the edges of the objects. Combining both the reconstruction from KinFu and the live point cloud results in a representation of the real-world space that is of greater quality that either can offer alone. Unfortunately, the KinFu algorithm will sometimes wrongly estimate the pose of the Kinect, causing a misalignment between the two models. Also, if the pose estimation fails completely, the algorithm will reset the reconstruction and it will need to be reinitialized. There are two main constrains to the use of the prototype imposed by the hardware. Firstly, the Kinect can only detect depth in the range of approximately 0.5 to 4 meters, which means that our prototype cannot be used for any application requiring users to manipulate objects with their hand. Secondly, the Wiimote has a field of view of 41° horizontally (4:3 ratio) which constrains how much the user can tilt the display before tracking is lost.

Despite all these issues, our prototype achieves the goal of being a user-perspective magic lens (as seen in Figure 6c), while also supporting a traditional device-perspective mode (Figure 6b). The user-perspective effects and other depth cues are all supported. With the frustum of the scene properly aligned with the real world, the apparent distance of objects is correct. Motion parallax is properly handled as the user moves either their head or the display. Occlusion is correct, with virtual objects properly occluded by the real objects that are in front of them. However, there are visible recon-

struction artifacts, particularly at viewing angles where the user is looking at an area outside the Kinect's current view. The robustness and speed of the reconstruction and the pose tracking also needs work; rapid movements can cause a loss of pose tracking or a reset of the KinFu reconstruction. Nevertheless, the end result is still quite compelling, as it offers a glimpse into the possible near future for hand-held AR interfaces.

We have implemented a user-perspective AR magic lens, and to our knowledge this is the first one of its kind. Unlike the common AR magic lens of the device-perspective type or hybrid systems such as [11], we cannot use video-see-through methods to achieve our desired effect and must reconstruct the world in real-time and render it from a different view. Although this naturally results in comparatively lower visual quality, our prototype still produced visibly acceptable results. While the limited depth range and tracking robustness does not yet allow for any truly practical applications, this prototype demonstrates that it may soon be possible. Better and more numerous depth sensors will reduce the depth range issue while increasing the resolution of the 3D model. We can also expect that hand-held devices will soon have the processing power needed to make an untethered version of this prototype possible. Overall, given the complexity of capturing and re-rendering user-perspective arbitrary scenes in real-time, the results we achieved with off-the-shelf hardware and software are very encouraging.

5 CONCLUSION AND FUTURE WORK

The main contributions of this work are twofold: First, we conducted a user study comparing user-perspective and device-perspective AR lenses on a selection task. We evaluated both types of lenses with phone-sized and tablet-sized form factors. Our results indicate that a tablet-sized magic lens is significantly better for selection in terms of performance time. A user-perspective magic lens provides a small significant benefit in terms of path distance necessary to perform selection tasks. Our qualitative results provided strong evidence that users may prefer a user-perspective magic lens for selection tasks, especially for large lens sizes, while a device-perspective lens is preferred for search tasks.

Our second contribution is a proof-of-concept user-perspective lens engineered using off-the-shelf hardware and software. We used a Kinect sensor, Wiimote, and open source software to create our prototype, which serves as a first implementation of this type of magic lens. We discussed the limitations of the current prototype and future possibilities.

For future work, we would like to investigate methods for reducing the overall weight and size of the next user-perspective magic lens. There are also additional features that could be added, such as a stereoscopic display for a truly natural view of the scene. Stereoscopy would also make transparent screens a viable option. As

previously mentioned, a transparent screen could cause issues with aligning the virtual annotations with the backdrop, when used as an AR magic lens. However, if a parallax barrier were added for an autostereo transparent display it would greatly reduce this problem. That would be a promising solution for a magic lens as as there would be no visual artifacts or errors with the real world content. These transparent AR displays would still require 3D reconstruction to do occlusion handling, so they would benefit from the same technology used by systems based on re-rendering.

REFERENCES

[1] S. Asteriadis, K. Karpouzis, and S. Kollias. Head pose estimation with one camera, in uncalibrated environments. In *Proceedings of the 2010 workshop on eye gaze in intelligent human machine interaction*, EGIHMI '10, pages 55–62, New York, NY, USA, 2010. ACM.

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

[3] R. T. Azuma. The challenge of making augmented reality work outdoors. In *Mixed Reality: Merging Real and Virtual*, pages 379–390. Springer-Verlag, 1999.

[4] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. Toolglass and magic lenses: the see-through interface. In *Proceedings of the 20th annual conference on computer graphics and interactive techniques*, SIGGRAPH '93, pages 73–80, New York, NY, USA, 1993. ACM.

[5] L. Brown and H. Hua. Magic lenses for augmented virtual environments. *Computer Graphics and Applications, IEEE*, 26(4):64 –73, july-aug. 2006.

[6] A. Butz, T. Hollerer, S. Feiner, B. MacIntyre, and C. Beshers. Enveloping users and computers in a collaborative 3D augmented reality. In *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, 1999. (IWAR '99)*, pages 35–44, 1999.

[7] O. Cakmakci and J. Rolland. Head-worn displays: a review. *Journal of Display Technology*, 2(3):199–216, sept. 2006.

[8] C. Coffin, C. Lee, and T. Hollerer. Evaluating the impact of recovery density on augmented reality tracking. In *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2011*, pages 93–101, oct. 2011.

[9] S. Feiner, B. MacIntyre, T. Hollerer, and A. Webster. A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. In *First International Symposium on Wearable Computers, 1997. Digest of Papers.*, pages 74–81, oct. 1997.

[10] J. L. Gabbard, J. E. Swan, II, and D. Hix. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence: Teleoperators and Virtual Environments*, 15(1):16–32, 2006.

[11] A. Hill, J. Schiefer, J. Wilson, B. Davidson, M. Gandy, and B. MacIntyre. Virtual transparency: introducing parallax view into video see-through AR. In *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2011*, pages 239–240, oct. 2011.

[12] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on user interface software and technology*, UIST '11, pages 559–568, New York, NY, USA, 2011. ACM.

[13] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ISMAR '07, pages 1–10, Washington, DC, USA, 2007. IEEE Computer Society.

[14] E. Kruijff, J. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *Proceedings of the 9th IEEE International Symposium on Mixed and Augmented Reality, 2010*, pages 3–12, oct. 2010.

[15] R. S. Laramee and C. Ware. Visual interference with a transparent head mounted display. In *CHI '01 extended abstracts on human factors in computing systems*, CHI EA '01, pages 323–324, New York, NY, USA, 2001. ACM.

[16] C. Lee, S. Bonebrake, T. Hollerer, and D. Bowman. A replication study testing the validity of AR simulation in VR for controlled experiments. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2009*, pages 203–204, oct. 2009.

[17] J. Looser, M. Billinghurst, R. Grasset, and A. Cockburn. An evaluation of virtual lenses for object selection in augmented reality. In *Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia*, GRAPHITE '07, pages 203–210, New York, NY, USA, 2007. ACM.

[18] J. Looser, R. Grasset, and M. Billinghurst. A 3D flexible and tangible magic lens in augmented reality. In *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007. ISMAR 2007.*, pages 51–54, nov. 2007.

[19] R. Newcombe and A. Davison. Live dense reconstruction with a single moving camera. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1498–1505, june 2010.

[20] R. Newcombe, S. Lovegrove, and A. Davison. DTAM: Dense tracking and mapping in real-time. In *2011 IEEE International Conference on Computer Vision (ICCV)*, pages 2320–2327, nov. 2011.

[21] J.-y. Oh and H. Hua. User evaluations on form factors of tangible magic lenses. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ISMAR '06, pages 23–32, Washington, DC, USA, 2006. IEEE Computer Society.

[22] T. Olsson and M. Salo. Online user survey on current mobile augmented reality applications. In *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2011*, pages 75–84, oct. 2011.

[23] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59:207–232, 2004.

[24] E. Ragan, C. Wilkes, D. A. Bowman, and T. Höllerer. Simulation of augmented reality systems in purely virtual environments. *IEEE Virtual Reality Conference*, 0:287–288, 2009.

[25] M. Rohs and A. Oulasvirta. Target acquisition with camera phones when used as magic lenses. In *Proceedings of the twenty-sixth annual SIGCHI conference on human factors in computing systems*, CHI '08, pages 1409–1418, New York, NY, USA, 2008. ACM.

[26] M. Rohs, A. Oulasvirta, and T. Suomalainen. Interaction with magic lenses: real-world validation of a Fitts' law model. In *Proceedings of the 2011 annual conference on human factors in computing systems*, CHI '11, pages 2725–2728, New York, NY, USA, 2011. ACM.

[27] D. Schmalstieg, L. M. Encarnação, and Z. Szalavári. Using transparent props for interaction with the virtual table. In *Proceedings of the 1999 symposium on Interactive 3D graphics*, I3D '99, pages 147–153, New York, NY, USA, 1999. ACM.

[28] A. State, K. P. Keller, and H. Fuchs. Simulation-based design and rapid prototyping of a parallax-free, orthoscopic video see-through head-mounted display. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '05, pages 28–31, Washington, DC, USA, 2005. IEEE Computer Society.

[29] A. Takagi, S. Yamazaki, Y. Saito, and N. Taniguchi. Development of a stereo video see-through HMD for AR systems. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality, 2000. (ISAR 2000).*, pages 68–77, 2000.

[30] J. Viega, M. J. Conway, G. Williams, and R. Pausch. 3D magic lenses. In *Proceedings of the 9th annual ACM symposium on user interface software and technology*, UIST '96, pages 51–58, New York, NY, USA, 1996. ACM.

[31] S. You and U. Neumann. Fusion of vision and gyro tracking for robust augmented reality registration. In *Proceedings of the Virtual Reality 2001 Conference (VR'01)*, VR '01, Washington, DC, USA, 2001. IEEE Computer Society.