8-15-2014

The Effect of an Occluder on the Accuracy of Depth Perception in Optical See-Through Augmented Reality

Chunya Hua

The effect of an occluder on the accuracy of depth perception in optical see-through augmented reality

By

Chunya Hua

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Computer Science
in the Department of Computer Science and Engineering

Mississippi State, Mississippi

August 2014

The effect of an occluder on the accuracy of depth perception in optical see-through augmented reality

By

Chunya Hua

Approved:

_____

J. Edward Swan II
(Major Professor)

_____

Cindy L. Bethel
(Committee Member)

_____

Stephen R. Ellis
(Committee Member)

_____

Edward B. Allen
(Graduate Coordinator)

_____

Jason M. Keith
Interim Dean
Bagley College of Engineering

Name: Chunya Hua

Date of Degree: August 15, 2014

Institution: Mississippi State University

Major Field: Computer Science

Major Professor: Dr. J. Edward Swan II

Title of Study: The effect of an occluder on the accuracy of depth perception in optical see-through augmented reality

Pages in Study: 63

Candidate for Degree of Master of Science

Three experiments were conducted to study the effect of an occluder on the accuracy of near-field depth perception in optical-see-through augmented reality (AR). The first experiment was a duplicate experiment of the one in Edwards et al. [2004]. We found more accurate results than Edwards et al.'s work and did not find the occluder's main effect or its two-way interaction effect with distance on the accuracy of observers' depth matching. The second experiment was an updated version of the first one using a within-subject design and a more accurate calibration method. The results were that errors ranged from –5 to 3 mm when the occluder was present, –3 to 2 mm when the occluder was absent, and observers judged the virtual object to be closer after the presentation of the occluder. The third experiment was conducted on three subjects who were depth perception researchers. The result showed significant individual effects.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my major professor Dr. J. Edward Swan II and my other graduate committee members. I would also like to thank everyone who helped me during my study at Mississippi State University.

This material is based upon work supported by the National Science Foundation, under awards IIS-0713609, IIS-1018413, and IIS-1320909 to J. E. Swan II.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	analysis of variance
AR	augmented reality
DLT	direct linear transformation
HMD	head-mounted display
IPD	inner pupillary distance
IRB	institutional review board
MAGI	microscope-assisted guided interventions
OST	optical see-through
VST	video see-through

CHAPTER I

INTRODUCTION

1.1 Development of AR

Augmented Reality (AR) is a promising technology which combines computer generated scenes with real environments in order to augment the physical world with useful information. According to Azuma [1997], an AR system should have the following three main characters:

"1) Combines real and virtual

2) Interactive in real time

3) Registered in 3-D"

The origin of AR can be traced back to the work of Sutherland [1968] in the 1960s. In his paper, a head-mounted 3D display was demonstrated, which was the initial demonstration of a head-mounted display. Since then, many applications of AR have been developed, especially in areas such as medicine, manufacturing, entertainment, military, etc. (Azuma [1997]). For example, many AR-assisted surgical systems have been developed and achieved some measure of success. Most of the surgical systems use an image for the augmenting information. AR-assisted surgical system usually provide medical images on patients' body parts that are hard to observe, for example, hidden capillaries and arteries, which can be virtually presented during operations. In manufacturing, AR technology can be applied to assist understanding, maintaining and

fixing complex machines. In addition, entertainment is another application area for AR technology. Here, the effort is on interactions between actors and virtual characters, and background enhancement in concerts, exhibitors, etc. Finally, AR is regarded as a next generation, promising method for military tasks. For example, head mounted displays and other wearable equipment have augmented the flight information given to pilots.

Looking back at recent developments in augmented reality, applications are becoming more mobile, interactive and commercial (Azuma et al. [2001]). With the rapid development and wide use of smart devices, various mobile augmented reality libraries have been produced, for example, ARToolkit [2014] for mobile platforms and Vuforia SDK [2014]. Most of the SDKs provide image-based tracking by taking advantage of computer vision technology. Also, with the interactive interfaces provided by tablet devices and phones, such as touch, slice, and zoom in/out, applications developed on these SDKs are interactive as well. In recent years, many companies, such as Total Immersion [2014], have opened up the commercial AR market. Recently many AR-assisted solutions have been introduced for different commercial bands. For example, in the Virtual Dressing Room Application [2014], customers choose and try on virtual clothes in front of a human-height display. The customer is tracked, and they can move and spin to view the virtual clothes on themselves.

1.2 Optical see-through AR VS video see-through AR

AR systems are usually classified as video see-trough (VST) AR and optical see-through (OST) AR according to the way users observe the real world (Azuma [1997]). A VST AR system usually contains a camera and a tracking system. The camera is normally used to capture real world scenes, while virtual objects are rendered into the

captured scenes to provide augmented information. The augmented scenes are then provided to the users by two possible ways. One is through a mobile tablet display such as a smartphone, and the other is through an immersive head-mounted display (HMD) where the surrounding environment is hidden. On the other hand, OST AR systems allow users to directly observe their surrounding physical space with their eyes. Virtual objects are rendered on an HMD display. Users see a combination of the real world and the virtual objects. In order to allow users to see both the real world and the virtual objects together, an OST HMD display needs to contain an optical combiner that is transmissive and reflective at the same time. One of the challenges of designing these displays is discovering the proper ratio of transmittance to reflection.

Both OST and VST systems have their own advantages and disadvantages. First, VST systems usually have more latency in blending than OST systems. This is because VST systems need cameras to capture the physical world, and then must render and align virtual objects with the physical environment, while OST systems only need to render virtual objects. In addition, the camera's update rate and the display's update rate both affect the final image update rate, which has a direct impact on the latency of VST systems. An image update rate of 60Hz means at least a delay of 16.67ms (Azuma [1997]). However, although VST systems have more latency in rendering than OST systems, it is easier for VST systems to compensate for the latency difference between the real and virtual channels. Because frames from either channel can be buffered until the frame from the other channel is available.

A second tradeoff is in resolution, In an OST system, the resolution of the real environment is not constrained. In contrast, with a VST system, because the real world is

displayed via captured video frames, the resolution is limited to that of the capturing camera.

A third tradeoff is in occlusion. Both OST and VST systems need depth information from the physical world and virtual world simultaneously, which is one of the biggest challenges for AR technology. For an indoor environment or a static environment, it is possible to create a depth map of the whole environment by simply manually reconstructing the 3D model. In addition, automatic 3D reconstruction for a static environment is also possible, from multiple images (Moons et al. [2010]) as well as from technologies like Kinect Fusion (Izadi [2011]). However, if an AR system is targeted to an outdoor or dynamic environment, creating and maintaining a depth map will be more challenging (Dou and Fuchs [2014]). 3D reconstruction technology based on one to several depth cameras is still believed to be a possible solution (e.g. Izadi [2011], Dou and Fuchs [2014]). Structured light has also used for depth extraction in an AR-assisted laparoscopic surgery application (Fuchs et al. [1998]). In addition to the depth map problem, true occlusion in VST systems is possible but not in OST systems. This is because OST systems suffer from two light source problems, one from the real world and the other from the virtual world. Imagine a scenario where a virtual cube is in front of a real sphere. Here, the cube should obscure the sphere. However, if no filter is applied to obscure the light coming from the real sphere, the user will see a mixed image of the sphere and the cube, which makes the cube look ghostly and transparent. However, the applied filter must not block the user's view, which is even more difficult. This leads to a tradeoff regarding how strong the light from the virtual world should be applied compared to the light in the real world, which is also known as the contrast problem.

Another tradeoff involves tracking. For VST systems the camera can supply another source of tracking information in addition to conventional trackers. Computer vision-based AR tracking technology has been developed and applied in libraries like ARToolkit [2014] and Vuforia SDK [2014]. With fiducial markers applied to the real world, computer vision algorithms can dynamically detect the location and rotation of the markers. This technology makes many AR systems conventional tracker-free, which has led to a thriving marketplace of AR applications on smart devices.

Lastly, how to achieve accurate registration is a problem for both systems. In VST system, accurate camera calibration and high quality tracking are two key factors. Camera calibration is a well-discussed topic in computer vision, and many algorithms have been developed for the popular pinhole camera model. For example, the direct linear transformation (DLT) method originally reported in Abdel-Aziz et al. [1971], the two-step Tsai's algorithm which solves the extrinsic and intrinsic matrices directly (Tsai [1986]), the flexible calibration method in Zhang's algorithm (Zhang [2000]), etc. High quality tracking can be achieved using marker detection from camera frames, as well as other technologies such as inertial-optical trackers, magnetic trackers, ultrasonic trackers, etc. Most of these trackers claim accurate tracking with errors smaller than several millimeters, but all have tradeoffs regarding tracking volume, latency, noise, and accuracy. In OST systems, tracking accuracy is constrained by the accuracy of HMD calibration, which is very difficult for many reasons, such as the shifting of the HMD on the head (Axholt [2011]). In order to reduce the difficulty level, in some cases, these movements are ignored and HMDs are calibrated with algorithms similar to those used in

camera calibration (e.g. Abdel-Aziz et al. [1971]). However, these methods have accuracy limits. Therefore, HMD calibration is still a difficult problem for AR.

1.3 Depth perception in AR

Depth perception in real world is well-studied. People use various depth cues for depth perception, which include but are not limited to an objects' relative size, familiar size, occlusion, texture gradient and height; perspective, accommodation, convergence and motion parallax (Cutting [1997], Cutting and Vishton [1995]). Based on distance, visual space is usually categorized as personal space which is within 2m, action space which is from 3m to 30m and vista space which is larger than 30m (Cutting and Vishton [1995]). From the perspective of the observer, distances can be classified as egocentric distances and exocentric distances. Egocentric distance refers to the distance from an observer to an object, while exocentric distance is the distance from one object to another object.

Depth perception in AR, on the contrary, is still an active research area where many phenomena and problems need to be studied. Also, with the development of AR technologies, especially the development of AR equipment, depth perception using developing equipment always needs to be tested. In addition to the conventional depth cues that work in real environments, researchers are eager to understand how other factors could affect the accuracy of depth perception. Many experiments have been done to understand how depth perception operates in AR environments, especially at medium-field and far-field distances (Bingham et al [2001], Creem-Regehr et al. [2005], Cutting [1997], Drascic and Milgram [1996]). Studying depth perception in AR is important for the development and improvement of AR-assisted applications; such as, for example,

AR-assisted surgical applications where the accuracy of depth perception is a critical factor. Using those AR-assisted surgical applications, surgeons usually need the ability to carefully put a physical tool, such as a needle, at a depth indicated by a virtual object.

Three major experiment techniques commonly used while studying depth perception are verbal report, perceptual matching and blind reaching. Verbal report is a cognitive method where observers report the distances they perceive to experimenters. However, with verbal reports the distance cannot be objectively measured in the real world, and there are concerns that verbal reports can be contaminated by cognitive knowledge that is unrelated to the perception of depth (Loomis and Knapp [2003]). Unlike verbal report, perceptual matching and blind reaching are not based on cognition, but are based on perception-action. Perceptual matching is a method where observers walk to a target object at medium/far distances, or reach out their hand to indicate the distance of a target object in near distances, while receiving visual feedback about their action. With blind reaching, on the contrary, observers walk or reach out their hands to a target object with their eyes closed. Therefore, the main difference between perceptual matching and blind reaching is whether or not observers receive visual feedback. Perceptual matching and blind reaching are more accurate than verbal reporting (Bridgeman [1997]).

1.3.1 Near Field Depth perception

In this work, we focus on near-field depth perception in AR environments. Unlike the large amount of depth perception experiments conducted to study medium and far field distances, to the best of our knowledge, there are only a few published experiments that study depth perception at near field distances in AR.

In Ellis et al.'s work [1998], they report four experiments that examine near filed depth perception in augmented reality using a see-through HMD. In their first experiment, they studied the effects of age, accommodative demand, and viewing condition, including monocular, biocular and stereoscopic viewing. The experimental results suggested that in the monocular viewing condition, when the accommodative demand was infinite (collimated), observer's AR perception was significantly biased towards the background. However, when accommodative demand was 50cm, young observers but not old ones perceived depth almost correctly in the monocular viewing condition, because young viewers are sensitive to accommodative cues. In the second experiment, they examined the effect of a physical surface, and found that the introduction of the surface could result in closer judgments. In their third experiment, Ellis et al. verified that the closer-judgment phenomenon in the previous experiment was related to a small convergence change after the appearance of the surface. In their fourth experiment, Ellis et al. suggested that the convergence change was more likely to be the cause of the closer judgments rather than the result of the occlusion itself.

Later, McCandless, Ellis, and Adelstein [2000] studied the effect of motion parallax and AR system latency on near field depth matching (75cm ~ 113cm), using the same apparatus as Ellis et al. [1998]. They studied AR system latencies of 31ms, 64ms, and 131ms, with monocular viewing. They found that errors in depth matching were significantly affected by the virtual target's distance, AR system latency, and also their interaction if motion parallax was applied. Without motion parallax, only the target's distance still affected depth matching accuracy (McCandless et al. [2000]).

Another study of near field AR depth matching is the doctoral thesis by Singh [2013]. In this work, the target's distance was in the range of 33 to 50cm. Singh examined two distance matching techniques: perceptual matching and blind reaching, and found that perceptual matching was more accurate than blind reaching. Singh also studied learning effects in his work, and found that learning effected matching in the real environment, but not in the AR environment. In addition, in another experiment Singh studied accommodative demand, including accommodation at an infinite distance (collimated), accommodation at the distance where the virtual object should be displayed (consistent), and accommodation at the middle of the distance range (midpoint). The results showed that accommodation did have significant effect on near field depth perception: errors decreased from 8~18mm to 2.5mm after changing accommodative demand from collimated to midpoint. Since people's accommodative ability strongly decreases with age, Singh also studied young and old observers. However, Singh did not find the expected result that young people would match the distance more accurately than older people. Instead, he found the opposite effect in the collimated condition and no differences between young and old observers in the other two conditions. Lastly, Singh also tested the brightness of the virtual object, and found that people observed the brighter virtual object to be closer than the dimmer one.

1.4 Medical AR development

In recent years, AR has been increasingly applied to medical applications. Medical AR applications usually involve image guidance and concern depth perception in the near filed, especially with AR-assisted surgical applications. Nowadays, AR systems have been proposed for neurosurgery (Kelly et al. [2013], Edwards et al. [2004]),

liver surgery (Scheuering et al. [2003], Suthau et al. [2002], Hansen et al. [2010]), craniofacial surgery (Tobias et al. [2002]), otolaryngology surgery (Edwards et al. [1995]), vertebroplasty (Bichlmeier et al. [2008]), laparoscopic surgery (Fuchs et al. [1998], Suzuki et al. [2008]), and microscope related surgery (Edwards et al. [1996], Johnson et al. [2001]).

Many obstacles needed to be overcome during the development of these medical AR applications. One of the biggest challenges involves the combination of medical and engineering expertise (Kersten-Oertel et al. [2013]). It is not hard to imagine that successful AR-assisted medical applications require successful collaborations between both medical and engineering scientists. However, most engineering scientists do not have a deep knowledge of the relevant medical domain, and vice versa, as medical scientists often do not have a strong engineering background. For engineering scientists, it is difficult to determine how to develop an AR application in a medical domain when medical scientists do not give critical but subtle requirement. The best situation would be if the engineering scientist happened to also be an expert in that medical area. Good communication is also very important. AR scientists and medical scientists should collaborate closely when developing the specific requirements of an application. In addition, consistent metrics for evaluating different solutions for AR medical applications should be collaboratively determined by scientists from both areas (Kersten-Oertel et al. [2013]). Therefore, multi-area experts and good collaboration are essential for developing useful and practical AR applications.

AR-assisted surgeries are also image-guided surgeries. Therefore, apart from difficulties attributable to AR specific techniques, AR-assisted surgeries also encounter

the same challenges as image-guided surgeries. The work of Kersten-Oertel et al. [2013] reviewed papers on image-guided surgeries from the point of data processing, visualization, and view. According to their work, there are two main kinds of data. One kind of data is patient-specific, like X-ray scanned data. The other kind of data is analytic which guides the surgeon with depth information and operation path information. The depth information usually refers to the depth of organs or the distance between hidden organs that are out of the surgeon's view, and the operating tools. This kind of information is typically indicated with numbers or decoded with colors. The operational path information refers to rendered data, like lines, that informs the surgeons about the possible operational direction and path of their tools.

Visualization difficulties include rendering, occlusion, registration, latency and many other problems. The focus of visualization studies is finding the proper depth cues for surgeons. The reported depth cues in Kersten-Oertel et al.'s [2013] are occlusion, the edges of virtual objects, textures with gradient information, stereoscopic view, motion parallax, shading, transparency, color-coding, and many others. The difficulty in achieving accurate near-field depth perception is also reported by others (e.g., Sielhorst et al. [2008]). Therefore, the medical AR communities agree that developing a better understanding of AR depth perception in the near field is a critical issue for AR medical applications. This is also the motivation behind this thesis work. In addition to choosing effective depth cues, accurate registration is also a large challenge. Registration mainly relies on accurate calibration, tracking techniques (Bajura and Neumann [1995]), and computer vision techniques for object extraction and tracking (Lepetit et al. [2003]).

View concerns first whether the patient or a monitor is the view location. In addition, Kersten-Oertel et al. [2013] considers various displays including HMDs, microscopes, stereoscopic displays, half-silvered mirrors combined with normal monitors, and using the patient as the image projection target. Finally, they consider interaction methods, including graphical user interfaces, radiologist assistance and conventional interaction tools like a mouse.

CHAPTER II

EXPERIMENT

As demonstrated in Chapter I, accurate AR depth perception is an open topic and it plays an import role in many AR applications. For example, in medical applications, surgeons usually need to operate by placing a surgical tool at a precise depth which is indicated by a virtual mark or indicator. However, to the best of our knowledge, only limited studies on depth perception in near-field Augmented Reality have been conducted, and even fewer have been targeted for medical AR applications. Apart from Ellis et al.[1998] and Singh [2013], Edwards et al. researched near-field depth perception in brain surgeries based on data collected from five surgeons (Edwards et al. [2004]).

Edwards et al. [2004] tested the microscope-assisted guided interventions (MAGI) system. This system was developed for AR-assisted neurosurgery and otolaryngology. During a surgery, the doctor could observe the patient's head, for example, in a craniotomy, through two lenses which combined the AR images from two monitors and presented this view to the doctor. The depth of the stereo image was determined by the coordinates of the left and right images displayed in the monitors. Also, the surgeon's inter-pupillary distance (IPD) was determined by conducting a special calibration step. During depth perception judgments, the subjects were asked to indicate the distance of the AR-presented virtual image information by placing a marker at the same distance. In their experiment, the virtual image which was a 2D green surface containing anatomical

data that could be displayed both over and underneath a plastic human brain model. Their first experiment studied the effect of the brain model's presence on the accuracy of depth judgments. It was based on data collected from 8 subjects. The results indicated that the presence of the brain model did have an effect on perceived depth, when the virtual image was displayed underneath the brain model. In their second experiment, they further studied this depth perception error, by showing the virtual image at 11 distances, both in front of and behind the plastic model. Five subjects were recruited in the second experiment. Figure 2.1 shows their result. For distances from 5mm to 20mm in front of the surface, depth perception was very accurate, with normalized error less than 0.5mm. However, for distances from 0mm to 20mm behind the surface, subjects overestimated with normalized error up to 2.5mm. Note that Edwards et al. [2004] normalized the error, in order to minimize any remaining IPD calibration error for each subject. The normalization method they used was to make the average error of the first four distances in front of the occluding surface be 0.

Our experiments were inspired by Edwards et al.'s [2004] work.

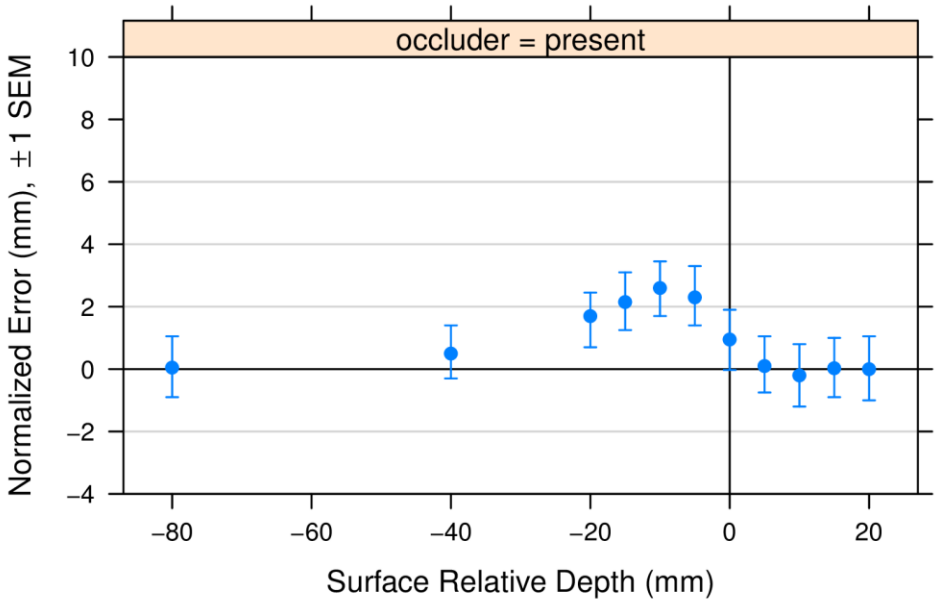

Figure 2.1 The result of Edwards et al.'s [2004] work

2.1 Experimental setup

2.1.1 Haploscope

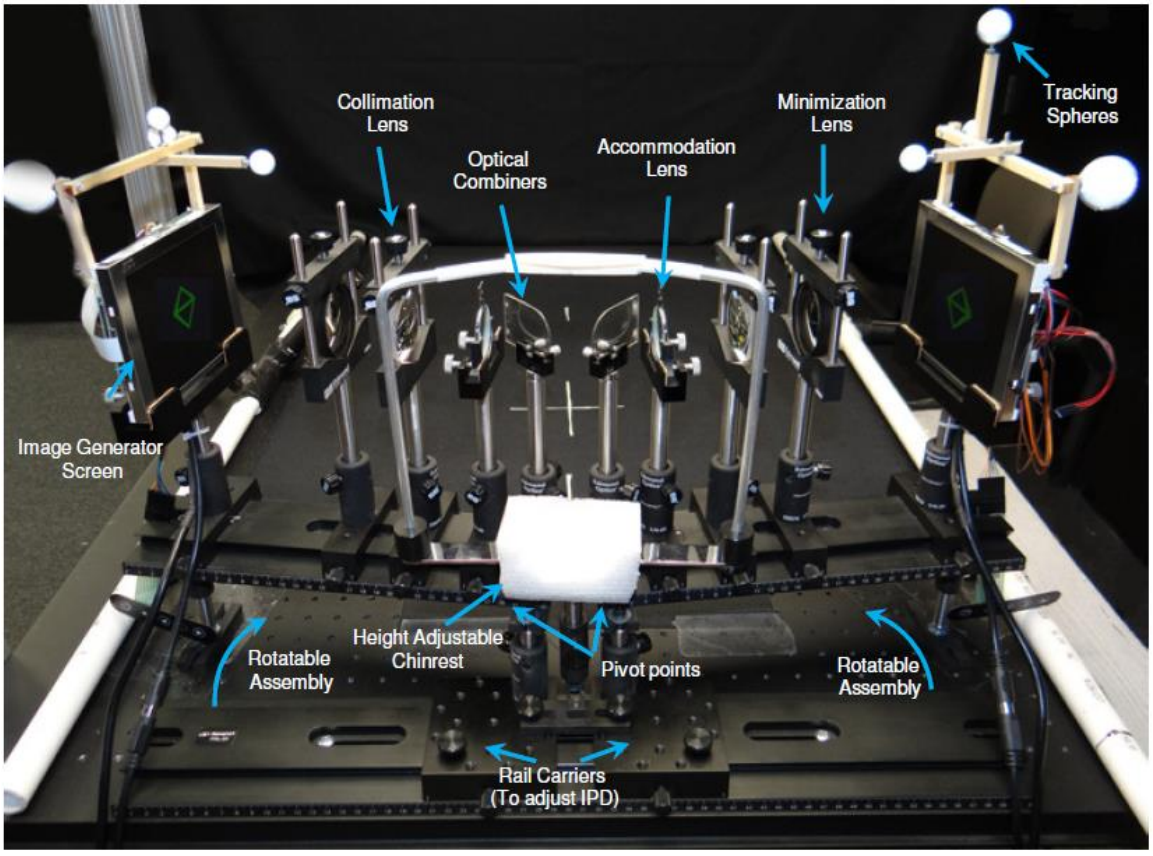

Figure 2.2 Haploscope system (Singh [2013])

Figure 2.2 shows our AR haploscope. The haploscope contains two commercial Accelevision LCD monitors, 9.2 cm in height and 12cm in width, which display virtual objects. Two concave lenses minimize the images in the monitors. Then, two convex lenses collimate the minimized images. Finally, another pair of changeable concave lenses allows the accommodative demand to be manipulated. Therefore, the haploscope allows us to independently manipulate accommodation and convergence angle. This

haploscope system was introduced in Singh's Ph.D. thesis [2013] for studying near-field depth perception in augmented reality. In the eyes, accommodation occurs when the eye focuses on an object. In these situations, at near-field distances, accommodation is a very important indirect depth cue, which affects the eyes' vergence angle (Ciuffreda [1992], Kirshnan et al. [1973] and Semlow et al. [1983]). In addition, Fisher et al. [1988] also found that accommodation itself can serve as a direct depth cue. Currently, many commercial HMD and other AR displays have a fixed focal distance, often infinity. The haploscope, on the contrary, can provide different focal distances by changing the accommodation lens. In addition, the haploscope allows us to study different vergence angles, through the rotation of whole light path from the monitors to the eyes, which allows the simulation of the vergence angles that occur when observers see a nearby object. And, unlike the real world, the vergence angle and the focal demand can either be consistent with each other, or they can diverge. This divergence allows us to study the effect of convergence-accommodation mismatches, which occur with any HMD that has a fixed focal distance. In our experiments, we focus on the study of depth perception at near-field distances, and the haploscope allows us to measure perceived distances with a large amount of accuracy.

2.1.2 Tracking system

A key part of the haploscope is the tracking system. In traditional AR systems, the tracking system is normally used to detect the transformation of the camera, physical objects, and users (Axholt [2011]). In this experiment, however, the tracker only detects the rotation of the two displays, and the location of the green LED light. The accuracy of the tracker has a direct impact on the accuracy of the haploscope system. Many

commercial trackers have been developed for VR, AR and motion capture applications. Popular trackers can be categorized as mechanical trackers, magnetic trackers, ultrasonic trackers, inertial trackers, optical trackers, and hybrid trackers. Axholt [2011] carefully reviews each kind of the tracker technology. Every tracker technology has its own advantages and disadvantages. For example, inertial trackers have very low latency, but even a small bias (around 0.001g) can accumulate into a large error (around 4.5m) in a short period of time (within 30s) (Axholt [2011]). On the other hand, inertial trackers are suitable for outdoor applications. Therefore, currently the most successful trackers are hybrid trackers, which combine two or more tracking technologies, and each technology's advantages can compensate for the other technologies' disadvantages.

The haploscope uses a SmartTrack infrared optical tracker (see Figure 2.3). Before tracking, the SmartTrack needs to calibrate its tracking space, and recognize tracking bodies consisting of arrangements of retro-reflective balls (see Figure 2.4). The tracking bodies are asymmetric, and therefore the SmartTrack can identify the way the bodies rotate and translate, and thus provide 6-degree-of-freedom data (the rotation as pitch, yaw, roll and the translation as x, y, z). The SmartTrack is very accurate. We ran the accuracy test shown in Figure 2.5. We measured 40 dots labeled on the desk 3 times each. The results were that the, SmartTrack's average 3D location error was within 2mm. In addition, the SmartTrack's rotation error was as high as $1/100^{th}$ of a degree (ART [2012]). However, the SmartTrack only focuses on a small tracking space of around $2m^3$. This is small compared to the large tracking space provided by some inertial-optical trackers such as the InterSense IS-1200. The SmartTrack produces data at a rate of 60fps, and it can simultaneously detect as many as four tracking bodies.

Figure 2.3 Infrared optical SmartTrack tracker

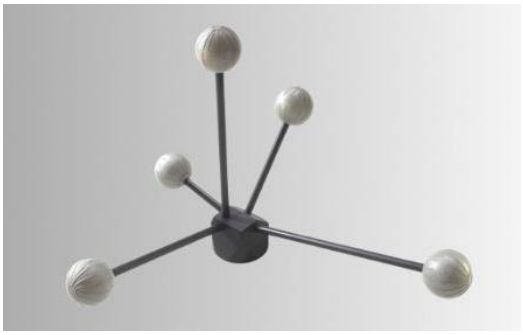

Figure 2.4 A SmartTrack tracking body

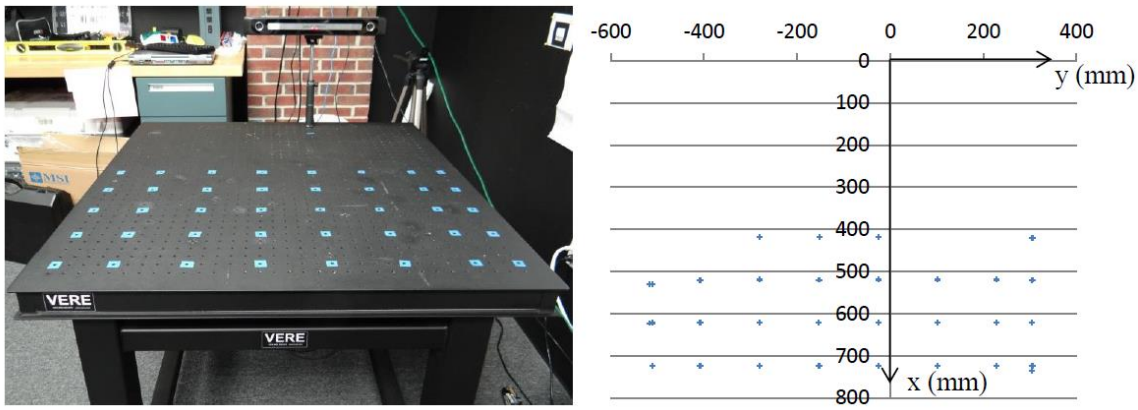

Figure 2.5 SmartTrack accuracy test

2.1.3 Using the SmartTrack with the Haploscope

As shown in Figure 2.2, we rigidly mounted two tracking bodies on top of the two monitors. The SmartTrack reports the rotation of the two tracking bodies. The yaw angle of the tracking body always equals the rotation angel of the corresponding monitor ($\alpha = \theta$ in Figure 2.6). Currently, all haploscope adjustments, such as rotating the two arms, manipulating the IPD, and changing the accommodation lenses, can only be done manually. However, a computer calculates the required yaw angle for the arms, based on the required distance of the virtual object. The geometry relationship among the angle θ, the distance d, and the IPD is :

$$\theta = \tan^{-1}\frac{IPD}{d} \tag{2.1}$$

Here, IPD is when an observer is converging at a distance of *d*. However, as shown in Figure 2.6, the observer's IPD is different when converging at different depths. The further the depth is, the larger the IPD becomes. In Experiment II, we measured each subject's IPD at different depths, and we found an average increase in IPD of 4.5mm when the vergence depth changed from 35cm to infinity. Based on Figure 2.6, this 4.5mm change could result in a small change of 0.34^{o} to 0.27^{o} in θ when d increases from 38cm to 48cm for a person with an average IPD range of 57.5mm to 62mm. In Experiments I and II, we used the IPD that we measured at optical infinity, and we did not change the IPD for different distances during these experiments. However, in Experiment III, we used the IPD that we measured at the distance of 400mm.

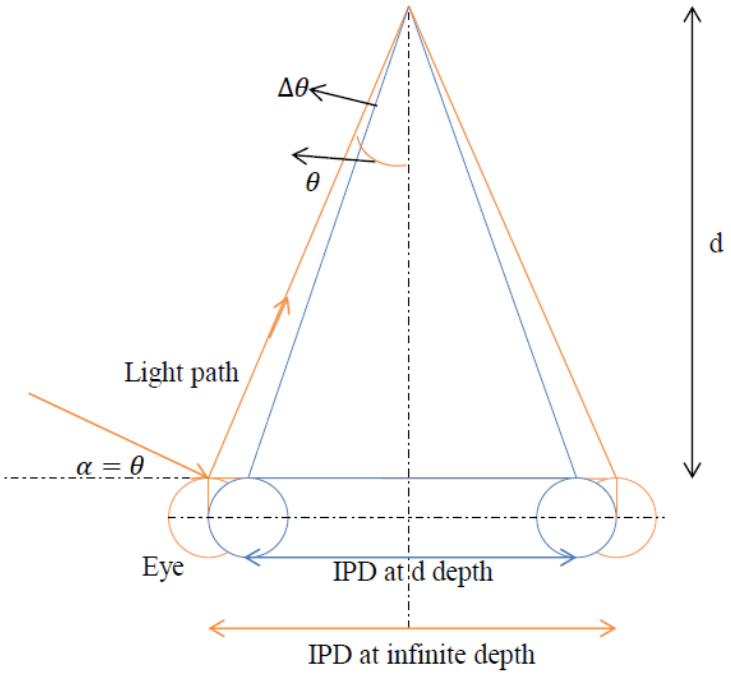

Figure 2.6 Geometry model of convergence, distance and IPD

2.2 Experiment I

2.2.1 Design

Experiment I aimed to find the effect of an occluding surface on the accuracy of near-field depth perception in Augmented Reality. The experiment was conducted with the haploscope system (Figures 2.7 and 2.8).

As shown in Figure 2.8, the system's environment was generally colored black. We covered the haploscope's table with a black cloth. In addition, we hung a black cloth in the background, in order to avoid conflicting depth cues, as well as to ensure that the green virtual target object was salient enough. In addition, monochromatic green light avoids chromatic aberration in the haploscope's optics, and green is the brightest color

that these monitors (and most monitors) can produce, because the human visual system is most sensitive to green wavelengths.

The experiment's primary variable was occluder status which includes occluder present and occluder absent conditions. For an occluder we chose a round disk, 30cm in diameter, printed with a highly salient black and white checkerboard pattern (Figure 2.7). We rigidly mounted it on a plastic pipe, which we could move to adjust its distance to the observer. This occluder was first introduced in Ellis et al. [1998]. The virtual object was a green octahedron, which was modeled after a physical octahedron that was 4cm in width, height, and depth (see Figure 2.7). Using the haploscope arms, we rotated the right and left images by a specific angle so that the observer could view the object naturally with 0 disparity (see Figures 2.12 and Figure 2.13, below). Both the occluder and the virtual object rotated at 4 rpm. We based this speed on the findings of Ellis et al. [1998]; they determined that the rotation speed of the occluder had no significant effect on depth judgment accuracy. Therefore, we chose a static rotation speed for both the occluder and the virtual object. We presented the occluder at a distance of 400mm in front of the observer, 26.7cm above the table surface. The virtual object appeared right above the lower rim of the occluder. Because the height of the virtual object in the physical world was static, at relatively farther distances, such as, for example, 480mm, the virtual object's vertical position was a little bit higher than the lower edge of the occluding disk. However, the virtual object always appeared horizontally centered on the occluding disk. During the experiment, we asked each subject to move the plastic pipe in their right hand forward and backward, which moved a small green physical LED light mounted on a pipe in depth. The subject's task was to align the LED with the lower point of the virtual

object. There was a small vertical gap of around 1.5cm between the LED and the virtual object's lower point. In some medical AR applications, especially in surgery applications, the depth indicator (the green LED tip in our experiment) should never touch the displayed virtual object. For example, in AR-assisted cardiac surgery, where the virtual objects are hidden arteries, veins, or capillaries, the surgeon's scalpels should never touch these important tissues.

For consistency, we chose the 11 distances that Edwards et al. [2004] also examined: 380, 385, 390, 395, 400, 405, 410, 415, 420, 440, and 480 mm from the observer. We presented the virtual target twice at each distance, but randomized the order so that subjects were not aware that distances repeated. However, we restricted this randomization so that each trial differed in depth by at least 20mm from the previous trial.

During Experiment I, we set the focal demand to 400mm, but provided an accurate vergence angle for each depth. Therefore, this experiment simulated a surgical application where the display had a fixed 400 mm focal demand, but vergence varied according to presented depth. Because AR displays do not typically have adjustable focal demand, this level of accommodation-vergence mismatch is what we might expect in a real near-field AR application.

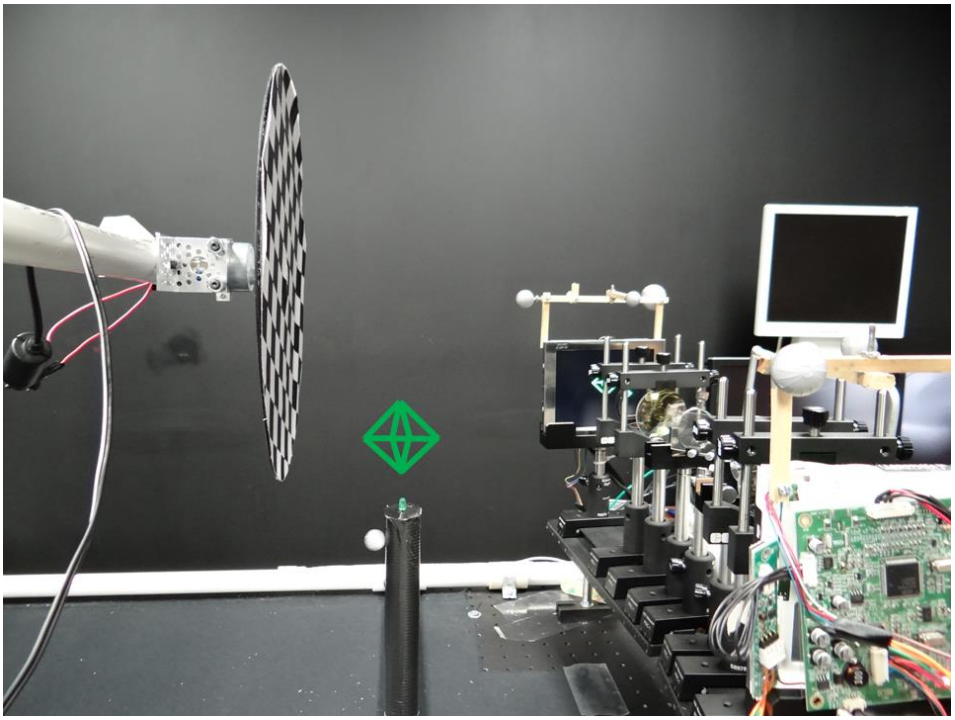

Figure 2.7 Experimental setup (side view)

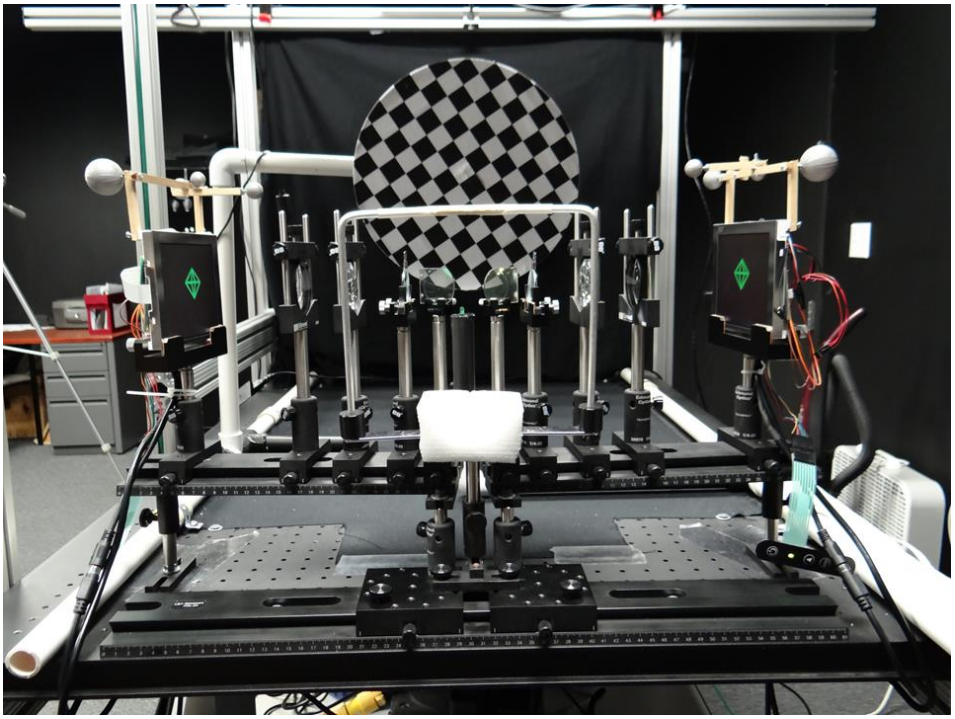

Figure 2.8 Experiment setup (front view)

We used a between-subject design for Experiment I. We recruited 22 subjects primarily from the student population of Mississippi State University. We reimbursed 7 subjects with course credit (s15, s16, s17, s18, s19, s20 and s21), and paid the remaining 15 of them $12 per hour. Because age affects accommodation ability, and because students are young, we only recruited young subjects. We considered four subjects to be pilot subjects, and used our experience with them to improve the experiment's final design. 9 subjects participated in the occluder = present condition, and 9 subjects participated in the occluder = absent condition. We later removed one subject's data in the occluder = present condition, and another subject's data in the occluder = absent condition, for reasons discussed further below. When choosing subjects, we did not consider gender to be an important factor, and the experimental subject recruitment system that we used at Mississippi State University concealed each subject's gender before they arrived for the experiment. However, in the end, in both occluder = present and occluder = absent conditions, 6 subjects were male and 2 female.

During pilot testing, we determined that subject were most comfortable completing 22 trials (2 repetition of 11 distances) within about 45 minutes, rather than completing 33 trials (3 repetition of 11 distances), which took a whole hour. The pilot data showed that subjects have the ability to carefully match the green LED tip vertically with the lower point of the virtual green rotating octahedron. Also, pilot subjects reported that they felt comfortable matching the indicator with the virtual object in a well-lit room, and therefore we kept the room lights on during Experiment I.

2.2.2 Calibration

Calibration is a critical procedure before conducting the experiment. Because different subjects typically have different IPDs, we conducted a calibration procedure to make sure that the virtual object was always displayed at the correct position for each subject.

As shown in Figure 2.9, we adjusted the haploscope IPD by moving the two rail carriers, and the true IPD is supposed to be the distance between the two stainless steel poles supporting the optical combiners, l. Our first calibration step was to set l to the subject's IPD, measured at optical infinity. In these photos, we used the IPD of 66mm, which is the author's IPD at optical infinity, for the human verification steps shown later. Our second step was to adjust the position of the virtual object's left and right images in the two monitors, in order to ensure that the haploscope correctly displays the images centered along parallel view vectors at an infinite distance. To achieve this, we created a calibration board, with two parallel vertical lines whose distance was equal to the subject's IPD at optical infinity. This calibration board could be placed at any distance in front of the subject along the center line of the haploscope system. Also, the two monitors must be in their initial positions, with rotation angles of zero (see Figure 2.9). After carefully adjusting the position of the virtual object, calibration circles in this step, each circle's center should coincide with the relevant calibration line drawn on the board (see Figure 2.10); this should align separately for both eyes. Because this calibration stage was set for optical infinity, the calibration results, which are the coordinates of the left and right images, should work for any IPD. In other words, because the monitors had zero rotation, the adjusted position of the virtual object in the monitors should work for

various IPDs, after carefully adjusting for l. After this step, the haploscope system was calibrated, and verified with naked eyes.

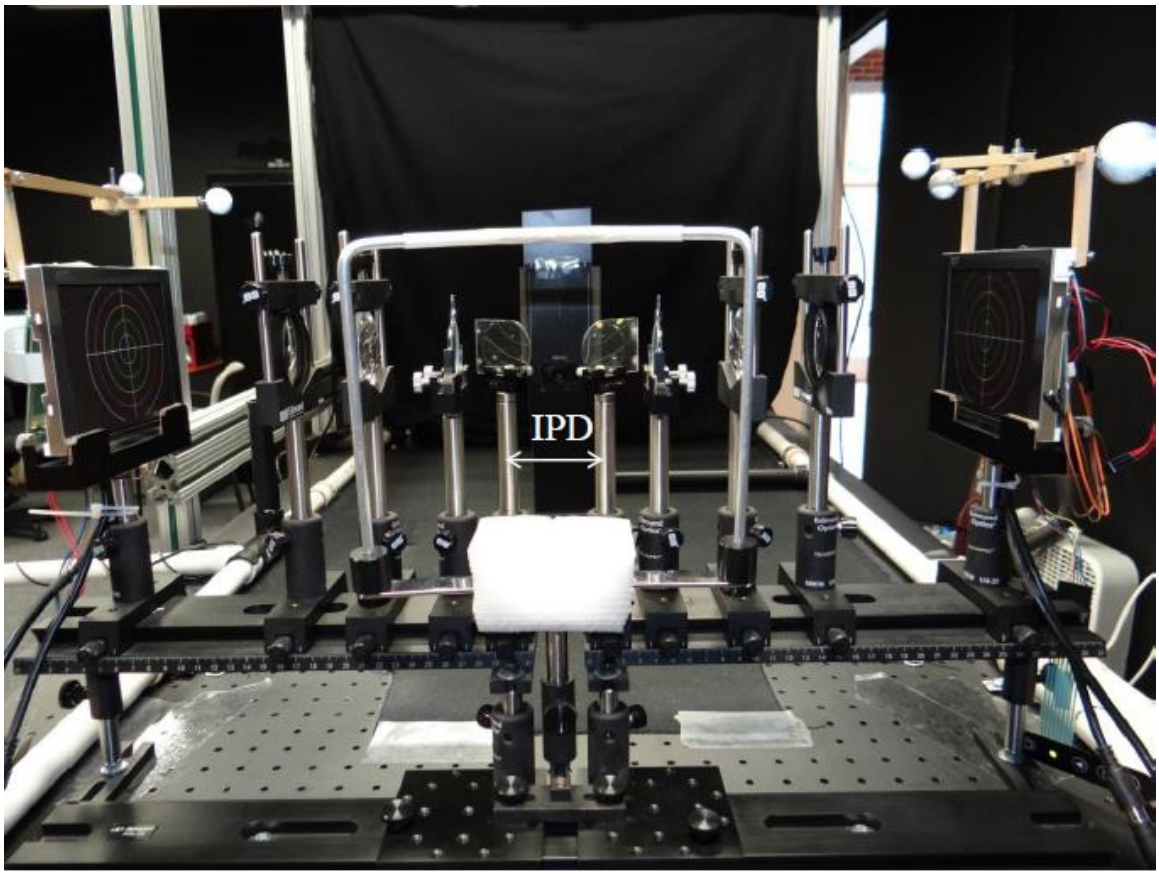

Figure 2.9 Calibration using naked eyes

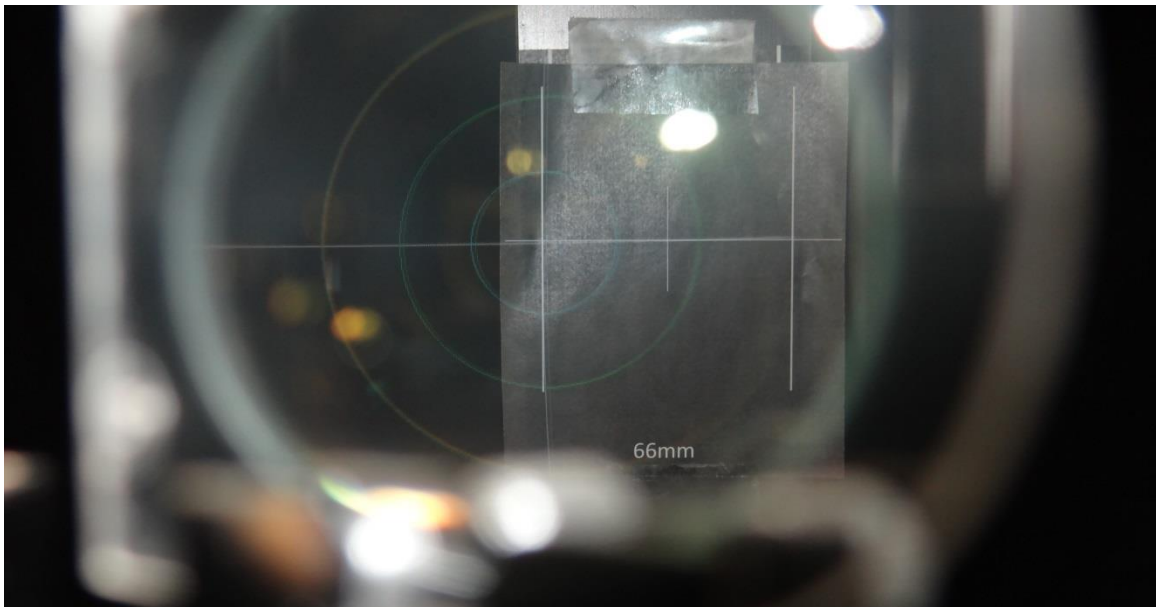

Figure 2.10 Calibration for the left image

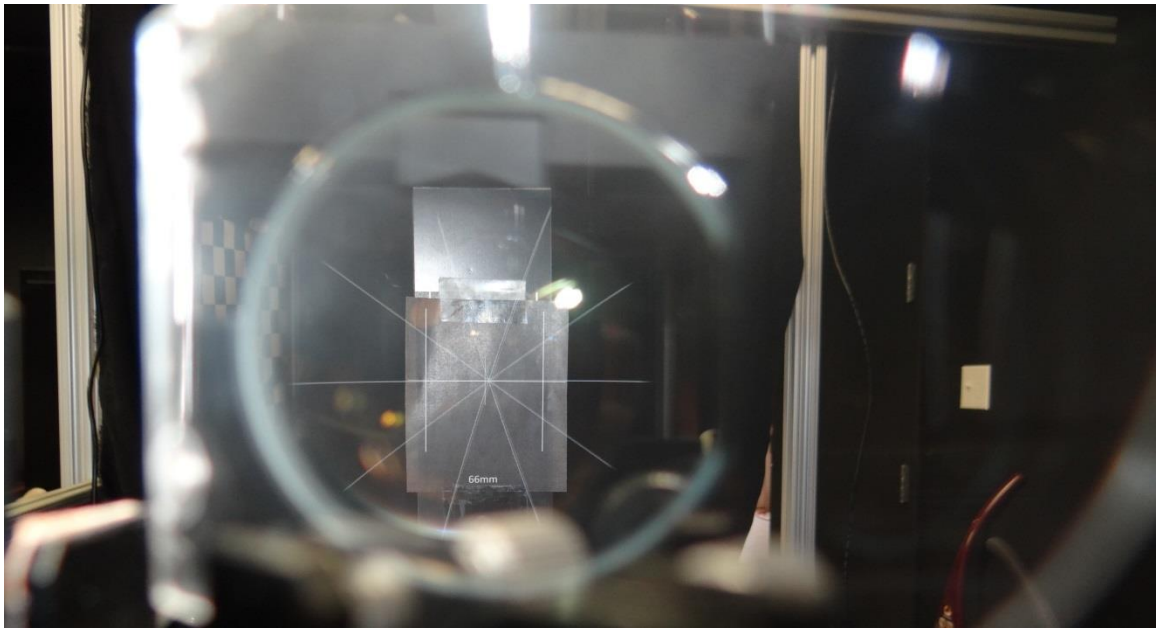

Figure 2.11 Verification using naked eyes

To verifying the calibration result, we tested the system at the distances that would be used in the experiment. As shown in Figure 2.11, the virtual object should appear right in the center of the calibration board, placed at a specific distance, after adjusting the rotation angle of the monitor according to the distance and the viewer's IPD. For additional verification, we hung a real object that had exactly the same shape as the virtual object in the true location where the virtual object should be displayed. We then mounted two Logitech cameras before the two optical combiners, separated according to the subject's IPD (see Figure 2.12). We examined all 11 distances that would be tested in our experiments, from 38cm to 48cm. If the calibration from the previous steps was correct, then the real and virtual objects should overlap in both camera views. Figure 2.13 demonstrates that the real and virtual overlaps were indeed very close.

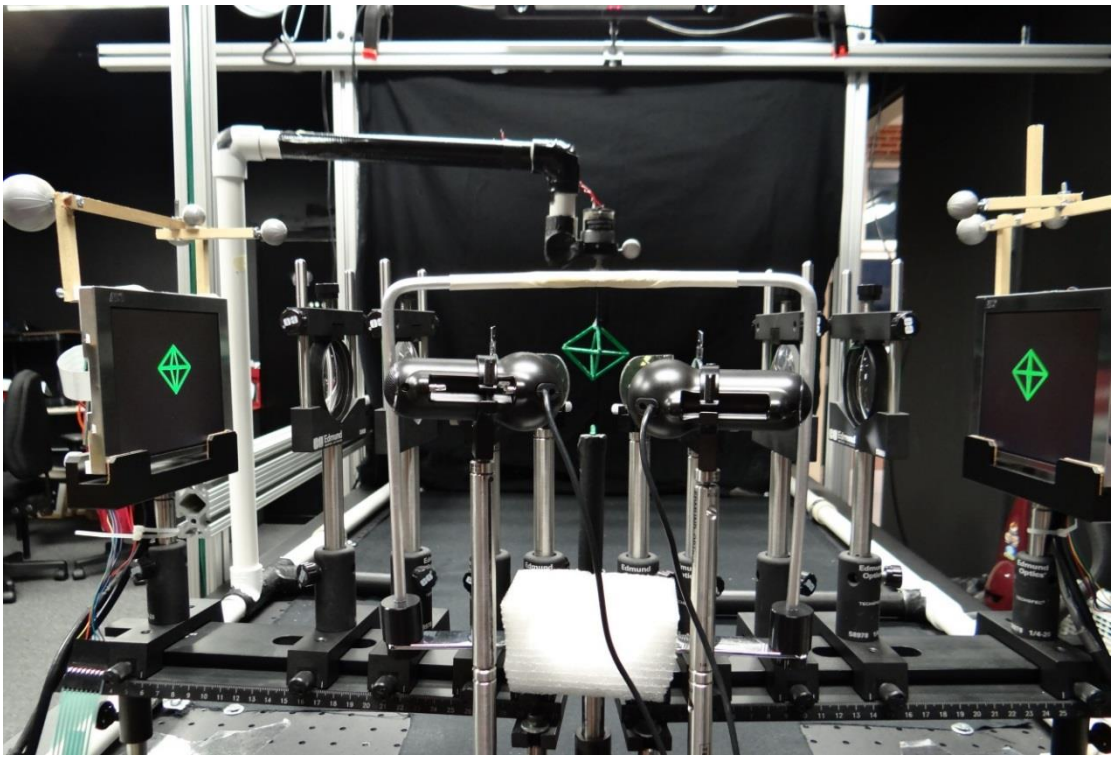

Figure 2.12 Verification using cameras

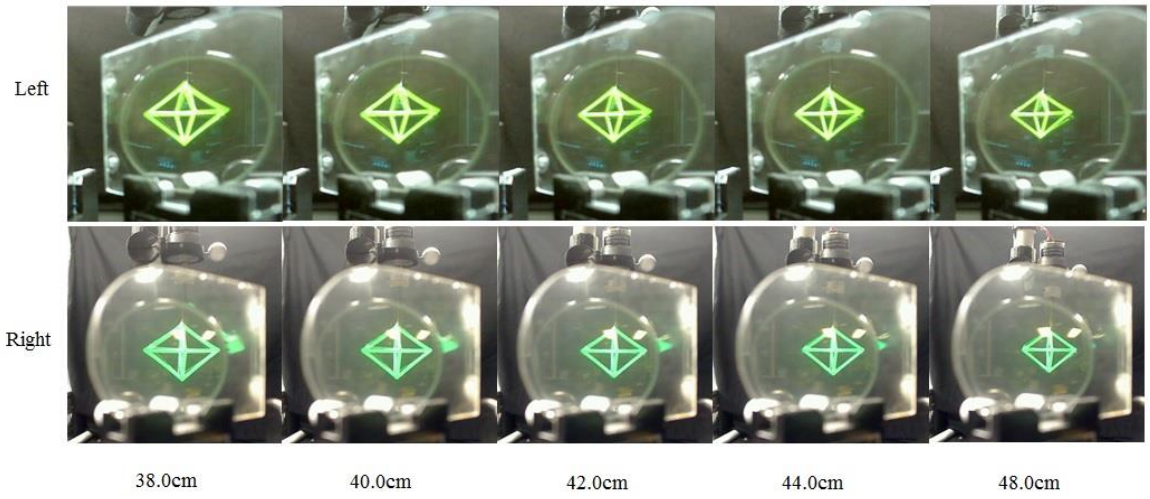

Figure 2.13 The result of verification using cameras

2.2.3 Procedure

Each subject first signed a consent form to indicate that they generally understood the experiment and its possible harm and benefits, meeting the requirements of Mississippi State University's institutional review board (IRB). Next, each subject completed a general questionnaire, which asked about the subject's familiarity with augmented reality/virtual reality/stereo glasses, their ability to mentally visualize and manipulate shapes or objects, their uncorrected vision, and the kind of glasses they normally wore. We encouraged each subject to wear any glasses they would typically wear when observing objects within their arm's length. In fact, all but one of the subjects who needed glasses at this distance wore their normal glasses; the one subject who did not reported that he was slightly nearsighted, but did not wear glasses in daily life.

Following the paper-work, we measured the subject's IPD, using a digital pupilometer. The pupilometer provided a digital readout for IPDs at different optical

distances from 30cm to 1m, and also at optical infinity. We only recorded the subject's IPD at optical infinity.

Before the main experiment conducted with the haploscope, we gave the subject a stereo vision test. This test showed 10 circles in stereo, 9 were at the same depth, and 1 was at a different depth. The subject's task was to identify which circle was at the different depth. As the test continued, the difficulty of the test increased: the depth of the target circle became closer to the other circles. This test measures the subject's ability to see increasingly fine stereo disparity differences. However, because the test uses red / blue anaglyphic stereo, it cannot be used for red / blue colorblind subjects.

Next, we set the haploscope to the subject's IPD. We only calibrated the system once, before running the first subject, using the method described above, and then left the system as it was. This could cause biased results, which we discuss later. After the IPD manipulation, we asked the subject to sit in front of the haploscope table and put their chain above a support frame. We next visually checked the subject's eye height and eye depth. The subject's eyes should be placed at the same height as the center of the combiner lenses; otherwise, the subject would see an optically distorted view of the virtual object, which could cause errors. The subject's eye depth was even more critical, because in the experiment, each subject measured the distance of the virtual object as it related to themselves: if the subject's eyes were placed a few millimeters farther than the right position, then they would report a slight underestimation in the matching task; similarly, if the subject's eyes were placed a few millimeters closer than the right position, then they would report slight overestimation in the matching task. Here, the right depth position for eyes was the zero position of the haploscope system.

After everything was prepared, we started the experiment. If the subject was recruited for the occluder = present condition, we placed the occluding surface 400mm in front of the subject. We displayed the virtual object at 11 distances with 2 repetitions. Before the very first trial, we moved the depth indicator to the closest position to the subject. For each trial, we asked the subject to carefully place the green LED directly under the bottom tip of the virtual rotating octahedron. Every time the subject finished a matching task, we asked them to say “ok” or “get data”, in order to inform the experimenter to record the position of the greed LED. We then continued with the next trial, until we completed all 22 trials. At any time during the experiment, the subject could withdraw unconditionally.

After the experiment, we encouraged subjects to report any discomfort caused by the experiment on a comfortableness questionnaire. Also, we asked subjects how confident they felt when they were matching the green LED indicator with the virtual rotating diamond. Finally, we recorded any reported difficulties in matching.

2.2.4 Result

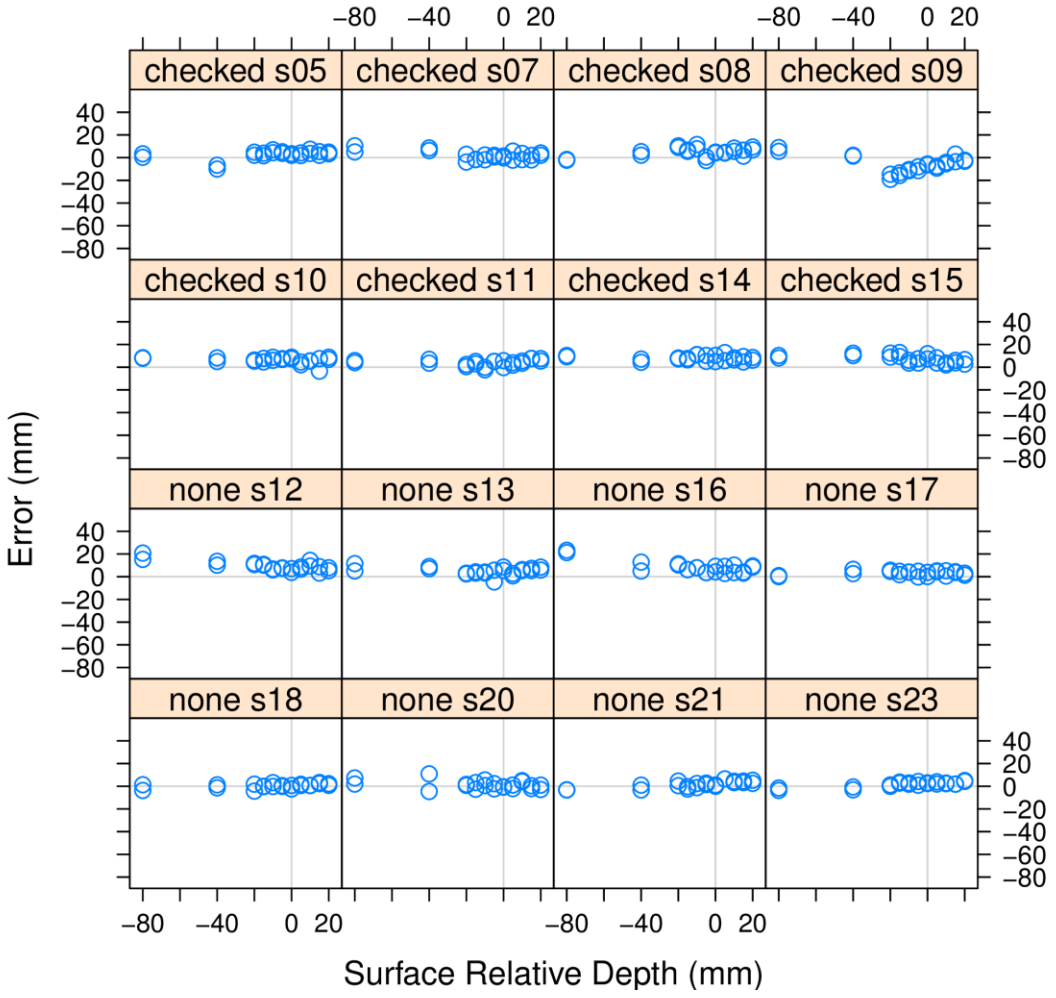

Figure 2.14 Error by each subject for Experiment I

8 subjects for occluder=present (checked) condition, and 8 subjects for occluder=absent (none) condition. Distances in front of the surface are positive values and distances behind the surface are negative values.

Figure 2.14 shows the error for each subject. We removed the data from subject s06 in the occluder = present condition, and the data of subject s19 in the occluder = absent condition. Subject s06 reported that they had a blurry view of the virtual target, and their data showed high variable error. Subject s19 got a low score in the stereo vision

test (38.46% correctness, compared to the average score of 76.44% for all other subjects), and they reported that they were not confident when matching. Their data was very different from the other subjects. Finally, subject s22 withdrew from the experiment due to equipment problems, but participated again the next day as subject s23.

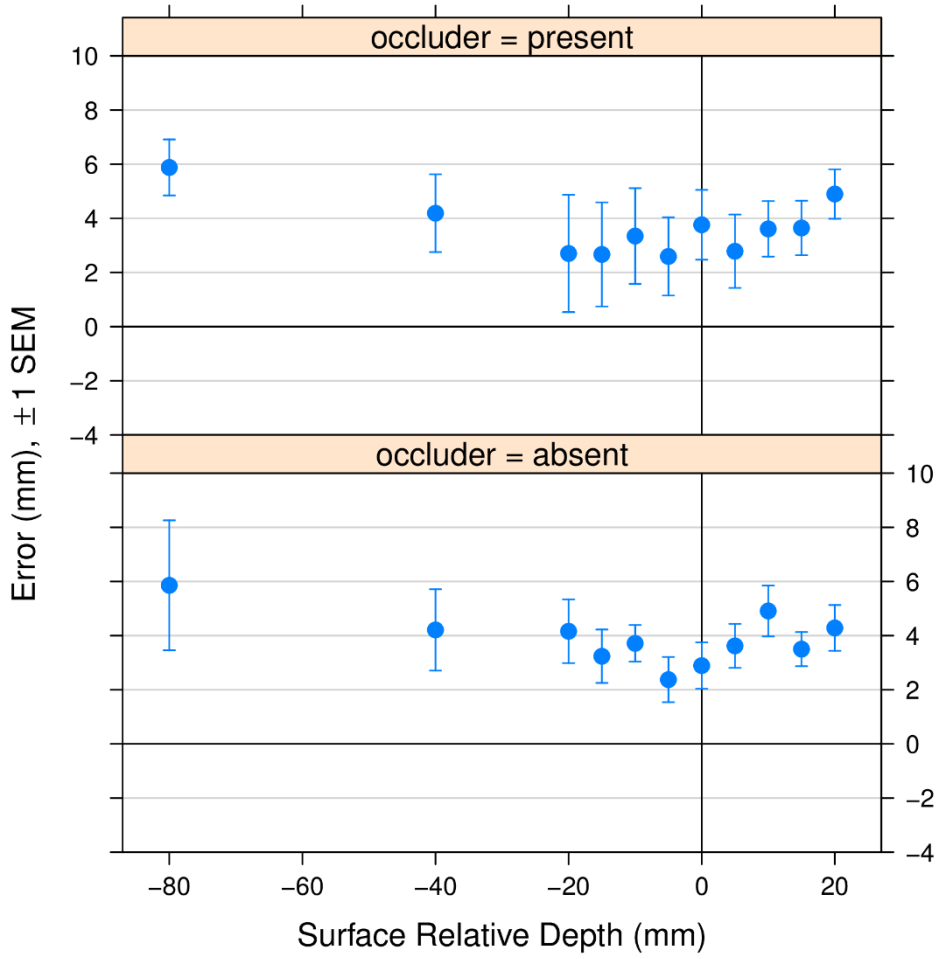

Figure 2.15 Error by occluder condition in Experiment I

Figure 2.15 shows the results for Experiment I. Overall, observers overestimated depth by 3mm to 6mm in the occluder = present condition and also in the occluder =

absent condition. A repeated-measures analysis of variance (ANOVA) shows that the occluder had no main effect on the accuracy of the participants' depth perception of the virtual target displayed at the 11 near field distances, $F(1,14) = 0.014$, $p = 0.91$. Also, the ANOVA showed that there was no occluder by distance interaction effect, $F(10,140) = 0.196$, $p = 0.996$.

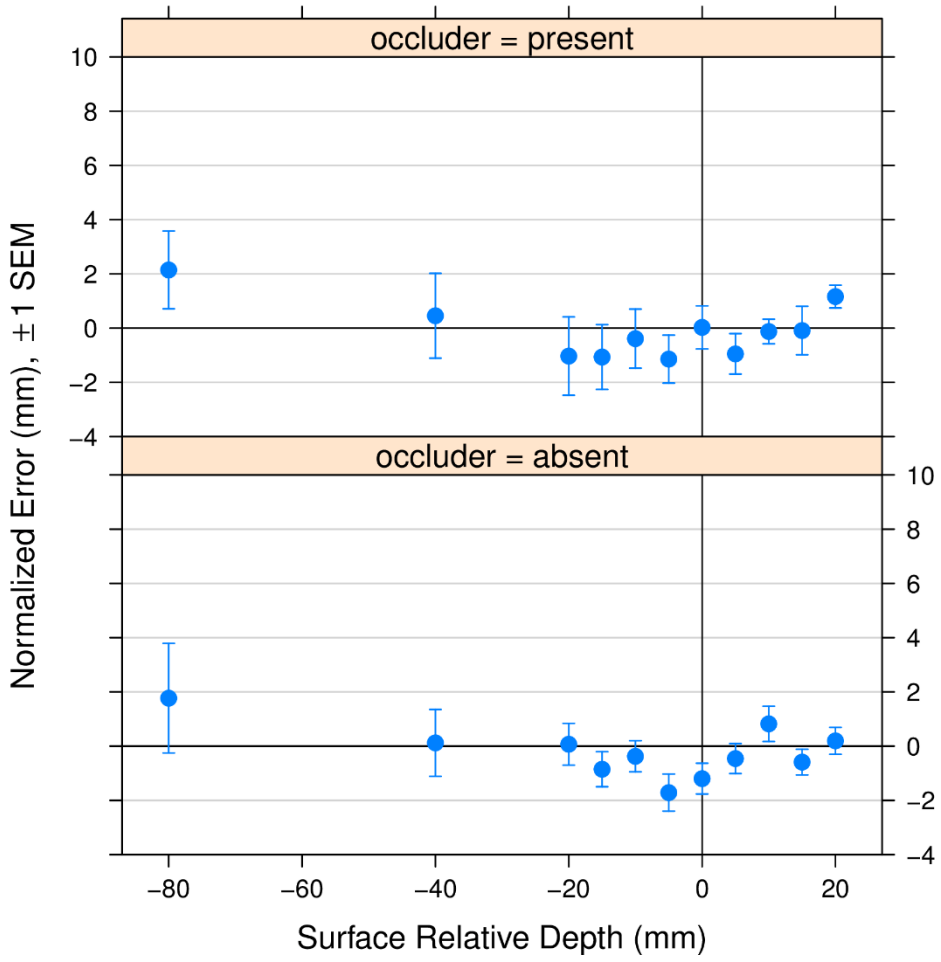

Figure 2.16 Normalized error by occluding condition in Experiment I

In Edwards et al.'s work [2004], they normalized their data by subtracting the mean error of the 4 distances in front of the occluding surface (see Figure 2.1). They did this in order to remove any remaining calibration error in their experiment. In order to compare our result with theirs, we also normalized our data in Figure 2.16, also by subtracting the mean error of the 4 distances in front of the occluding surface. The results are that observers had underestimation of less than 1mm in the first 20mm behind the occluding surface. We ran a repeated-measures ANOVA on normalized error, and did not find a main effect of occluder, $F(1,14) = 0.014$, $p = 0.906$, nor did we find an occluder by distance interaction effect, $F(10,140) = 0.196$, $p = 0.996$. This result was consistent with the results before the errors were normalized.

Figure 2.17 analyzes the coefficient of variation (SD/M) across the subjects. SD/M was larger when the occluder was present for the distances around the occluder (20mm to 5mm). Although a repeated-measures ANOVA showed that the occluder condition had no main effect on the coefficient of variation, $F(1,14) = 0.321$, $p = 0.58$, there was a marginal occluder by distance interaction effect, $F(10,140) = 1.701$, $p = 0.086$.

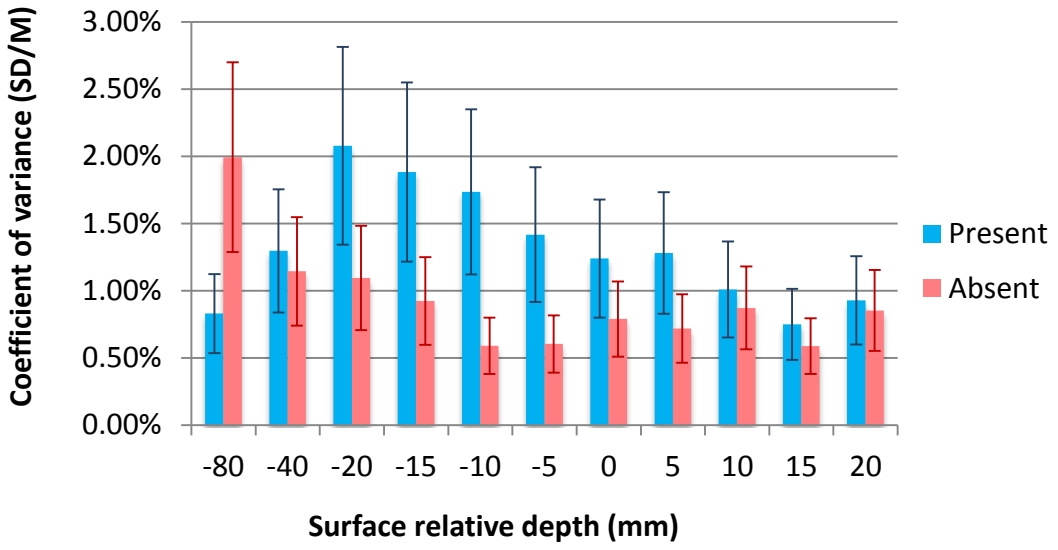

Figure 2.17 Coefficient of variation (SD/M) across subjects by occluder condition in Experiment I

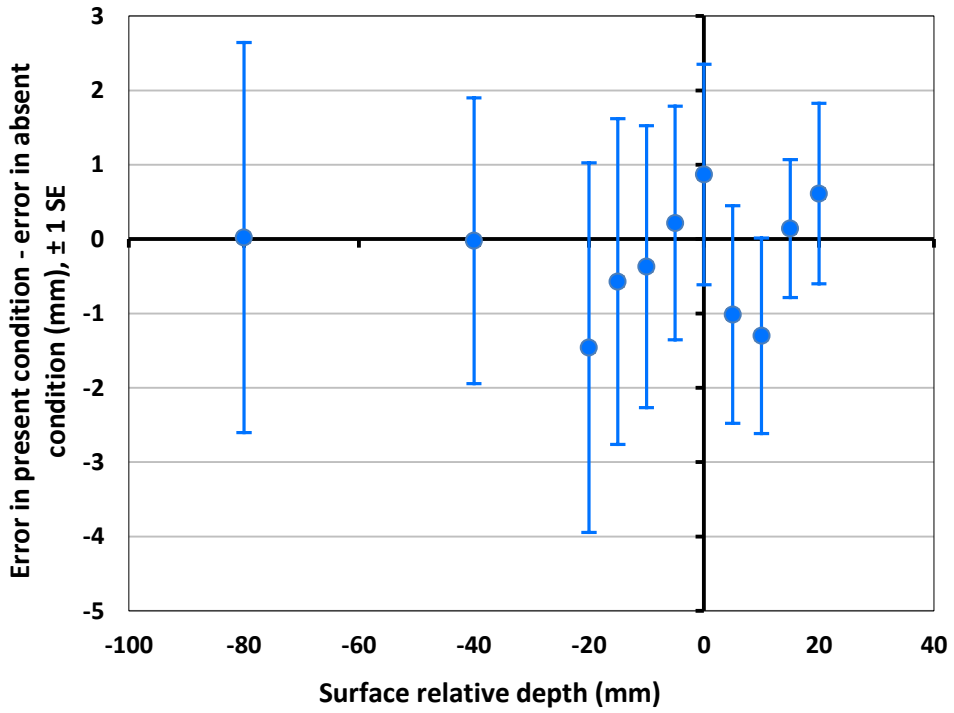

Figure 2.18 The difference between the errors in occluder=present condition and the errors in occluder=absent condition.

Figure 2.18 examines the difference between the errors in the occluder = present condition and the errors in occluder = absent condition. For the y axis, negative values mean that subjects observed the virtual object closer after the introduction of the occluder, which we call a push-forward effect; and positive values mean a push-backwards effect. We found an effect of the occluder on the judged distance only when the virtual target was displayed near the occluder, from –20mm to 20mm. In this range the observers tended to judge the distance of the virtual object closer, except for the three distances –5mm behind the occluder, the same distance as the occluder, and 20mm in front of the occluder.

2.2.5 Discussion

Comparing our normalized results (Figure 2.16) with the results in Edwards et al. [2004], overall, our results were more accurate. In Edwards et al.'s result, they found an overestimation of $2.1 \pm 3\ mm$ in the first 20mm behind their occluding surface, which was a plastic brain model. In our results, we found a small underestimation of less than 1mm in the first 20mm behind the occluder. One possible reason why we achieved more accurate results was that we used a more salient occluding surface, and therefore it was easier for observers to identify the virtual object against the occluding surface. Another possible reason was that our haploscope could supply more accurate stereo depth cues, compared with the system used by Edwards et al. [2004]. As previously discussed, our haploscope provides independent vergence and accommodation demands, while Edwards et al. only adjusted the coordinates of the images in their monitors to supply vergence.

With the presentation of the occluder, the ANOVA analysis showed neither a main effect of the occluder nor an interaction effect with distance on the accuracy of depth perception (Figure 2.15). The difference between the errors in the occluder = present condition and the errors in the occluder = absent condition (Figure 2.18) also confirms this result. For distances around the occluder, subjects observed the virtual target closer at distances of –20mm, –15mm, –10mm, 5mm, and 10mm, but further at distances of –5mm, 0mm, 15mm, and 20mm, when occluder was presented compared to when the occluder was absent. Overall, the effect after introducing the occluder was noisy. Also, the variation in the error difference was large. We believe that the main reason for these findings was because this was a between-subject experiment, and the effect sizes seem to be smaller than individual differences. This motivates the within-subjects design of Experiment II.

Another important aspect of the experiment that we care about is how accurate the results could get. How accurate should the depth judgments be in real medical applications? The answer depends. In highly precise surgeries like brain surgeries and cardiac surgeries, it is more likely that surgeons' manipulation need to be extremely accurate. In these scenarios, depth perception in AR-assisted medical applications must also be extremely accurate. In other scenarios, for example, depth perception in an AR-assisted application for simple skin suturing, the task may tolerate an error of several millimeters. Currently, we want to achieve the highest possible accuracy. The result in Figure 2.15, however, suggests that there was probably a bias in the data, especially under the assumption of accurate haploscope calibration. We suspect that the calibration was not as accurate as we intended, and therefore the calibration error resulted in a

convergence angle that was too small, which further resulted in a bias of seeing the virtual object as being too far. Therefore, in Experiment II, in addition to a within-subject design, we improved our calibration method, and we subsequently anticipated that the bias would be reduced or removed from the results.

2.3 Experiment II

Experiment II was an improved version of Experiment I. In this experiment, we used an improved laser calibration method and a within-subject design. The hypothesis was that there would be no or at least less calibration error in the data, and therefore the results would have more validity.

2.3.1 Design

The design of Experiment II was the same as Experiment I, except for what is discussed here. In Experiment II, we recruited 10 subjects. 8 of the subjects also participated in Experiment I. Here, we are interested in achieving the greatest possible accuracy, so we valued the practice that the subjects had with the task in Experiment I, and furthermore the task is low-level and psychophysical in nature, so subjects' prior knowledge of the task or purposes of the experiments is unlikely to affect their accuracy. In Experiment II we used a within-subjects design, and therefore each subject took part in both occluder conditions. A within-subject design factors away individual differences between subjects; in effect each subject acts as their own control. Furthermore, the within-subjects design additionally increases the practice that the 10 subjects had with the task. Each subject completed two sessions on two separate days, where in one session they performed the occluder = present condition, and in the next the occluder = absent

condition. We permuted the order of condition presentation (present, absent or absent, present) between subjects. Unlike Experiment I, in Experiment II each subject completed 33 trials for each condition (3 repetitions of 11 distances).

2.3.2 Calibration

As discussed above, we felt that the calibration method used in Experiment I was not optimal. One reason is that the video verification method takes a long time, and therefore it is not possible to complete the verification steps within a timeframe (5–10 minutes) that makes sense for a subject who expects to be done within an hour. Also, since the experimenter could only calibrate the system with their own IPD, for subjects with a different IPD, the haploscope had to be adjusted manually. However, if there is no verification step involved, the manual IPD adjustment could introduce unexpected errors, which we believed occurred in Experiment I. Therefore, we developed the following laser calibration method. We used two lasers to simulate the path that light follows to the subject's eyes (see Figure 2.19). Two laser levels were mounted behind the two optical combiners; the distance between the lasers was equal to the calibrating IPD. Each laser produced a vertical fan of light. Once the laser fan reached the optical combiners, the light reflected and reached the monitors following the reverse path that light emitted by the monitors follows to reach the subject's eyes. In addition, the laser light was also transmitted through the combiners, and reached the calibration board; the combiner was thin enough that the offset of the laser after the transmission could be ignored. For one calibration stage the two light fans remained parallel, and the transmitted light overlapped the two vertical lines on the calibration board (Figure 2.20x). In another calibration stage we canted the laser levels so that the two light fans converged to a specific near-field

distance (Figure 2.20y); here, the transmitted light combined and reached the central line instead.

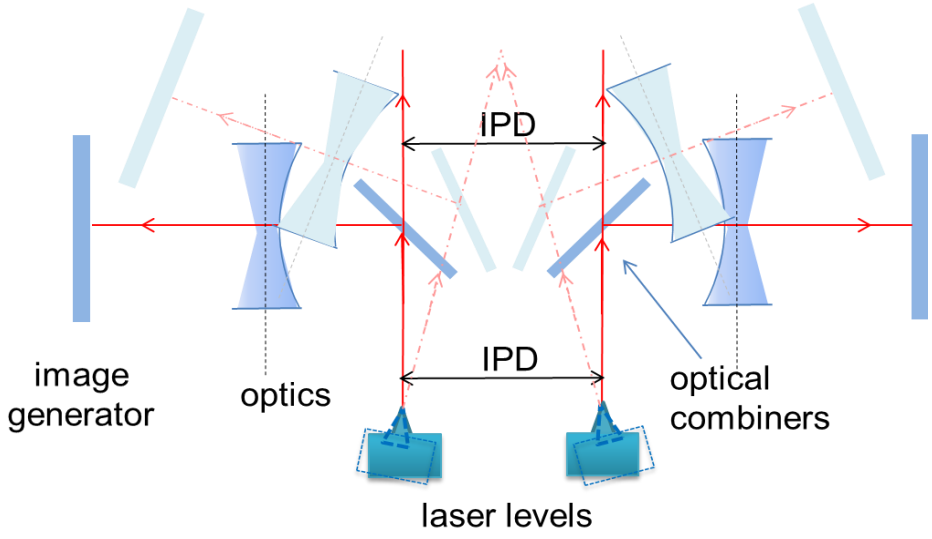

Figure 2.19 Top-down view of the laser-based IPD calibration method

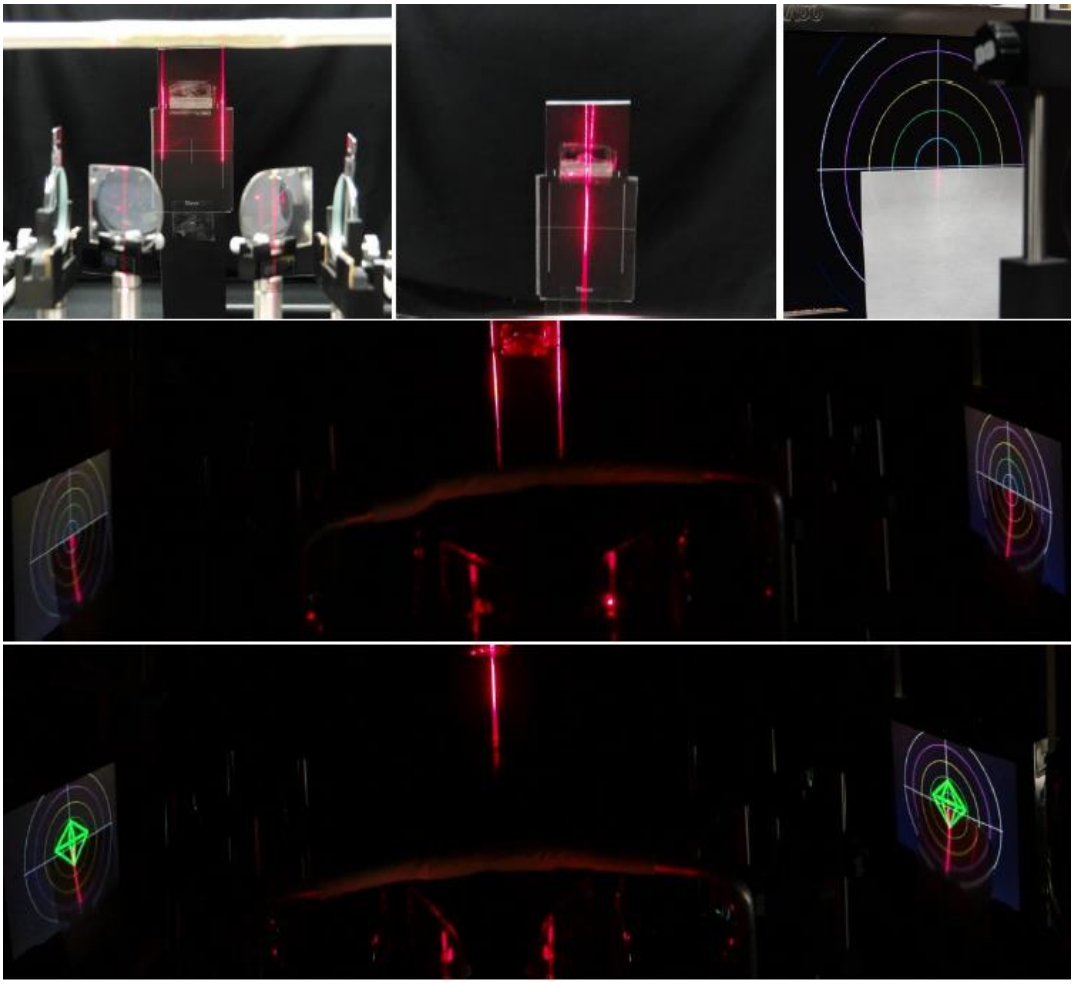

Figure 2.20 Laser calibration method

2.3.3 Procedure

The procedure of Experiment II was similar to that of Experiment I, except that the experiment included two sessions for each subject.

2.3.4 Result

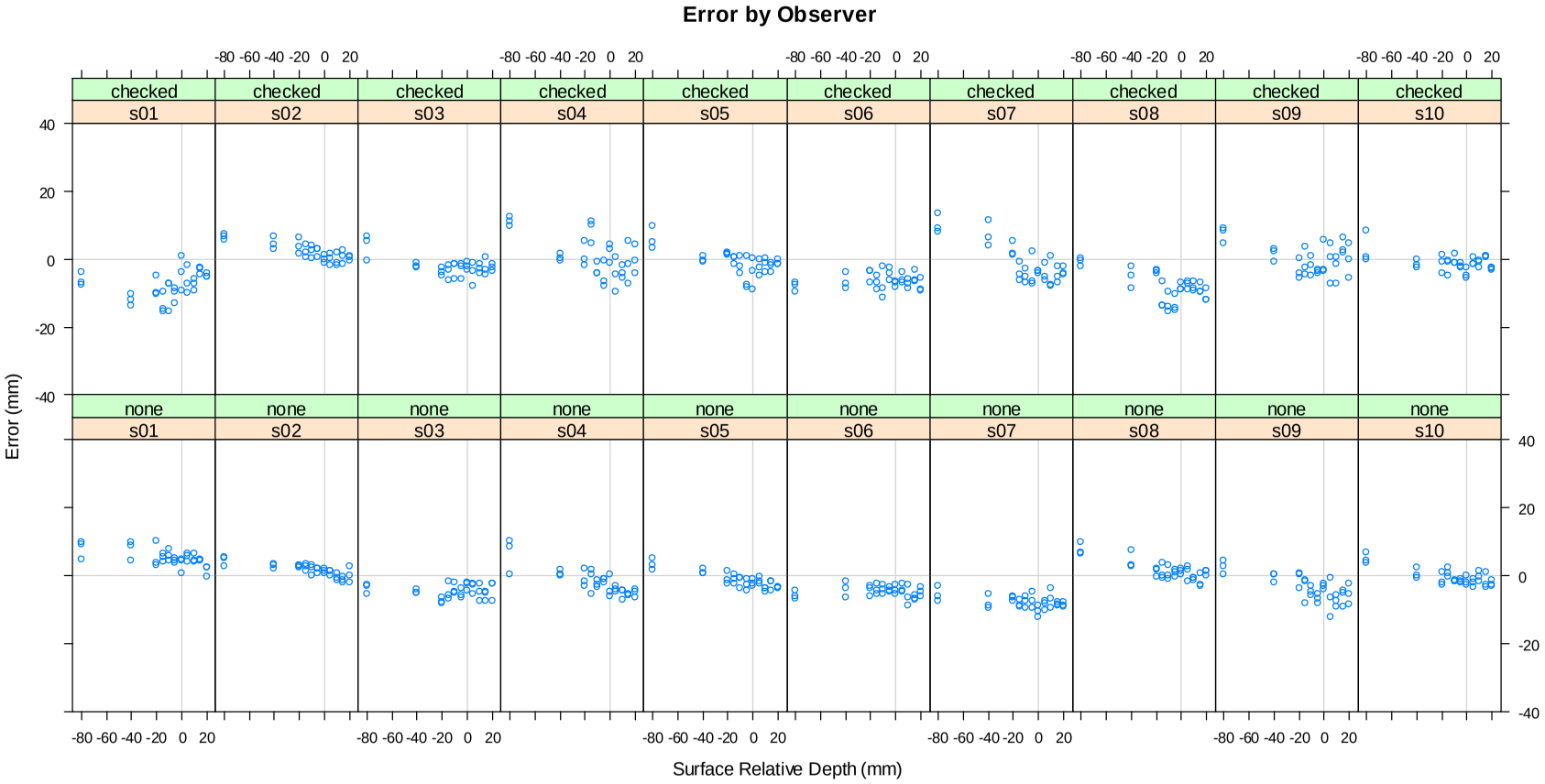

Figure 2.21 Error by observer in Experiment II

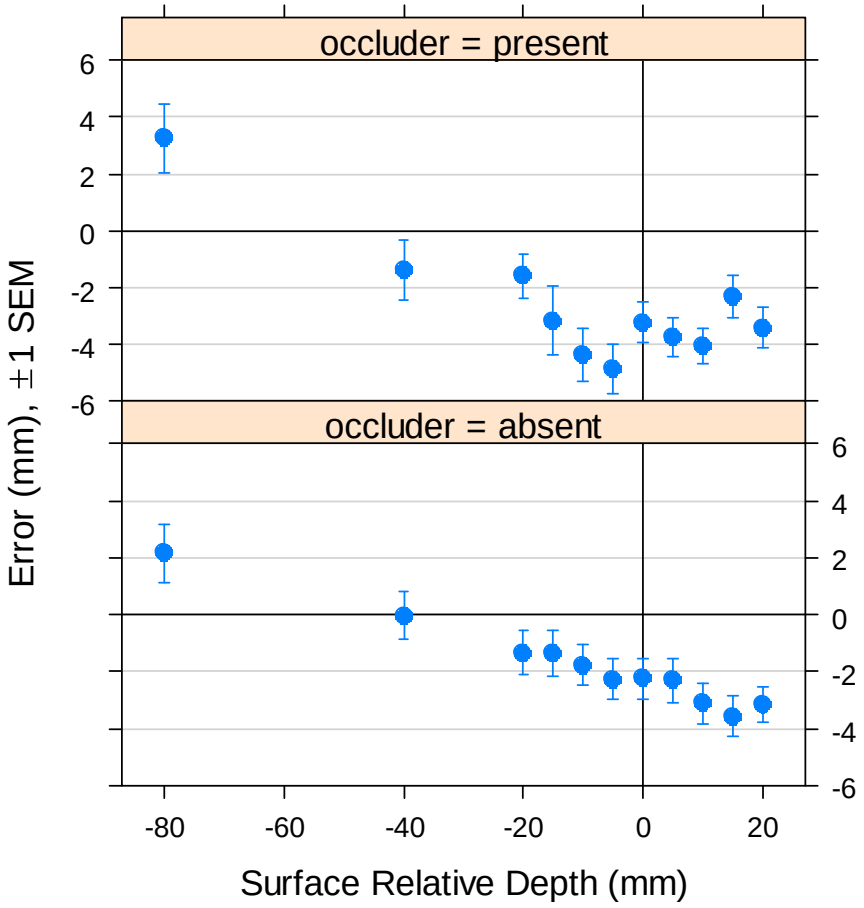

Figure 2.22 Error by occluder condition in Experiment II

Figure 2.21 shows the error by observer. Overall, the precision of the data in the occluder = present condition was not as good as when occluder was absent.

Figure 2.22 shows the averaged error by occluder condition. We found underestimation in both conditions, and the error was larger when the occluder was present: observers underestimated depth by 1mm to 5mm at distances from 20mm to –40mm, and only overestimated depth by 3.5mm at the distance of –80mm. However, without the occluder, observers underestimated 1mm to 3.5mm at distances from 20mm to –20mm, and overestimated by 2mm at the distance of –80mm.

A repeated-measures ANOVA revealed that there was no main effect of the occluder on error, $F(1,9) = 0.216$, $p = 0.653$. However, and different from Experiment I, there was a significant interaction between occluder and distance, $F(10,90) = 2.047$, $p = 0.037$*. This interaction is caused by the different pattern of results by distance for the occluder conditions (Figure 2.22).

The coefficient of variation (SD/M) on errors across the subjects was stable when the occluder was absent but not when occluder was present (see Figure 2.23). A repeated-measures ANOVA showed that there was no main effect of the occluder on the coefficient of variation, $F(1,14) = 0.451$, $p = 0.51$. However, within the subjects, the coefficient of variation on errors was larger when the occluder was present (see Figure 2.24). Here, a repeated-measures ANOVA showed a main effect of occluder, $F(1,9) = 36.18$, $p < 0.01$**, but no interaction of occluder by distance, $F(10,90) = 0.927$, $p = 0.513$.

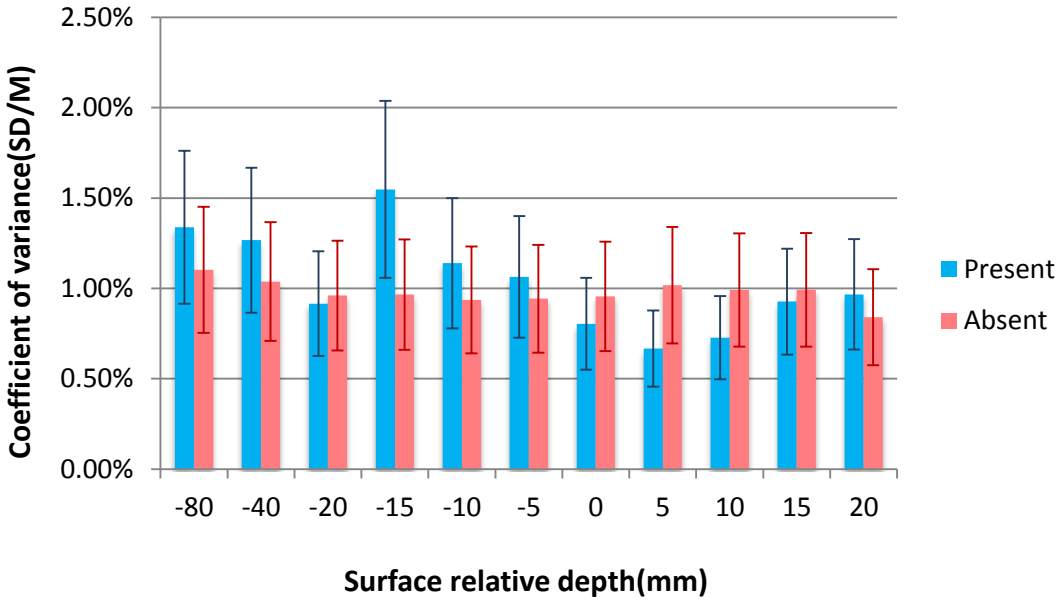

Figure 2.23 Coefficient of variation (SD/M) across subjects by occluder condition in Experiment II

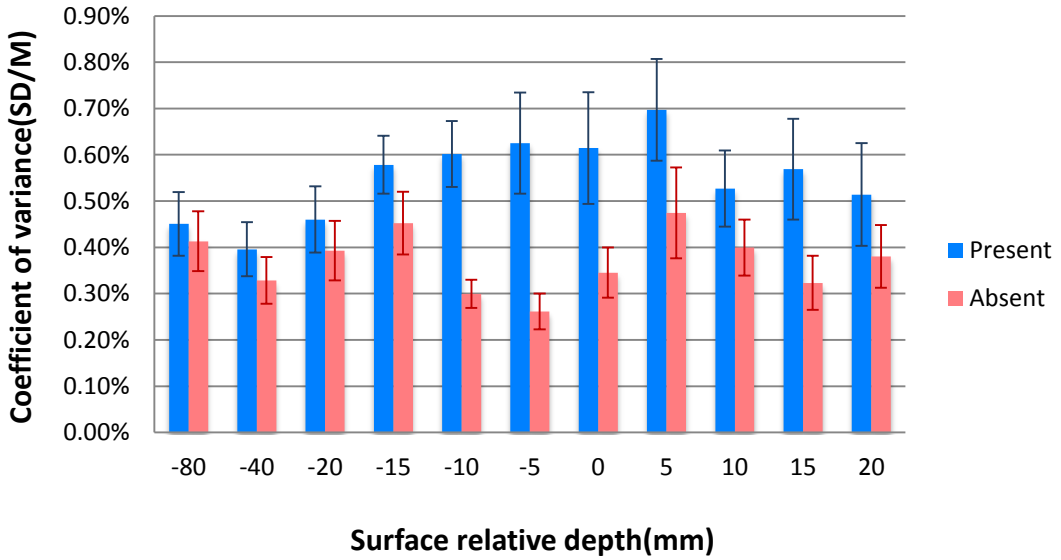

Figure 2.24 Coefficient of variation (SD/M) within subjects by occluder condition in Experiment II

2.3.5 Discussion

According to Figure 2.22, in the occluder = absent condition, the underestimation decreased as depth increased; it reached 0mm at the depth of –40mm and became an overestimation at the depth of –80mm. This was an interesting phenomenon. As we know, an observer's IPD becomes smaller when looking at a nearby object, because the observer's eyes need to converge to focus on near targets. We recorded each subject's near IPD from 35cm to 50cm and also their IPD at optical infinity, and found that there was an average change of 4.5mm in IPD. Because, in this experiment, we applied the IPD at optical infinity both in the software and in the haploscope, and if we only consider the effect of the IPD changing to converge on nearby objects, then we can expect some degree of underestimation. According to Figure 2.25, the real convergence for the virtual object at distance d should be α', which is larger than the α convergence that we used. Therefore, the subject underestimated the virtual object at a distance less than d. In addition, as d becomes larger, the IPD approaches the IPD at optical infinity, and α' approaches α, which explains why the underestimation error decreases with increasing distance.

In addition to the effect of the IPD change, accommodation can also affect depth judgments. In this experiment, we set accommodation to 400mm; and therefore, if we only consider the effect of accommodation, and assuming that all subjects have good accommodation ability, then in the occluder = absent condition, overestimation will be excepted for distances in front of 400mm and underestimation will be excepted for distances behind 400mm. However, as can be seen in Figure 2.22, our results are not consistent with this hypothesis.

Considering the effect of the IPD change and accommodation together, and also considering our results in the occluder = absent condition, it seems that the effect of the IPD change—the convergence effect—overcame the effect of accommodation. However, in order to replicate and validate this hypothesis, as future work we need to conduct another experiment where we change the IPD according to each distance in the software. If the error of depth perception in the occluder = absent condition becomes smaller, or the result is an overestimation to underestimation trend with increasing distance, then we will have verified that the effect of IPD change is significant in near field depth perception.

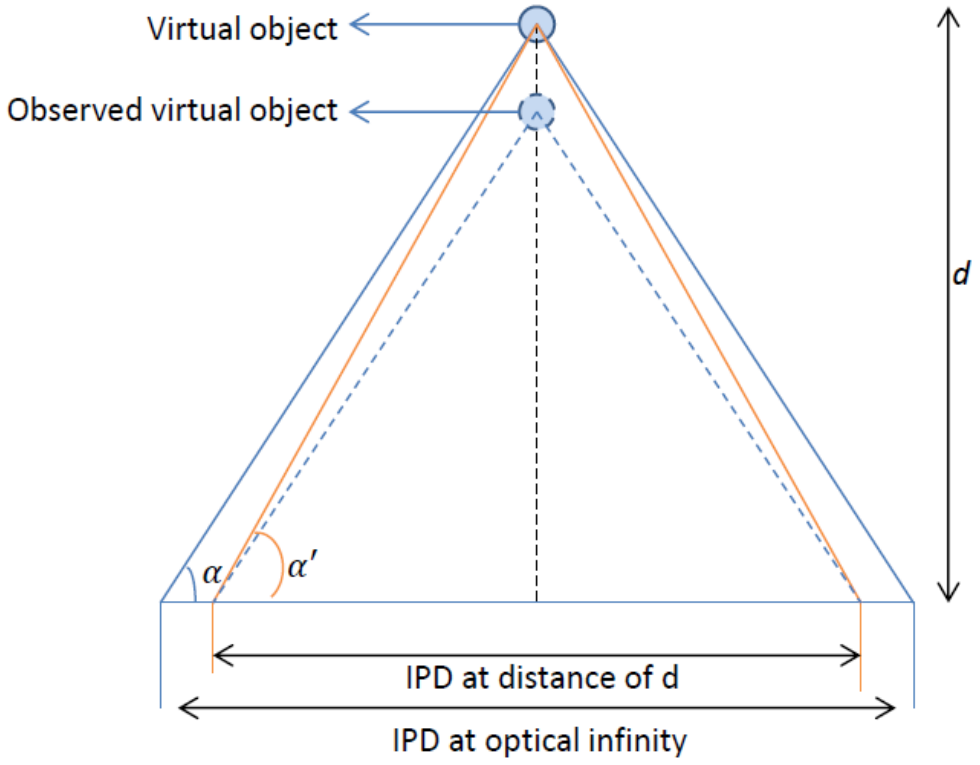

Figure 2.25 Near IPD effect on observed distance

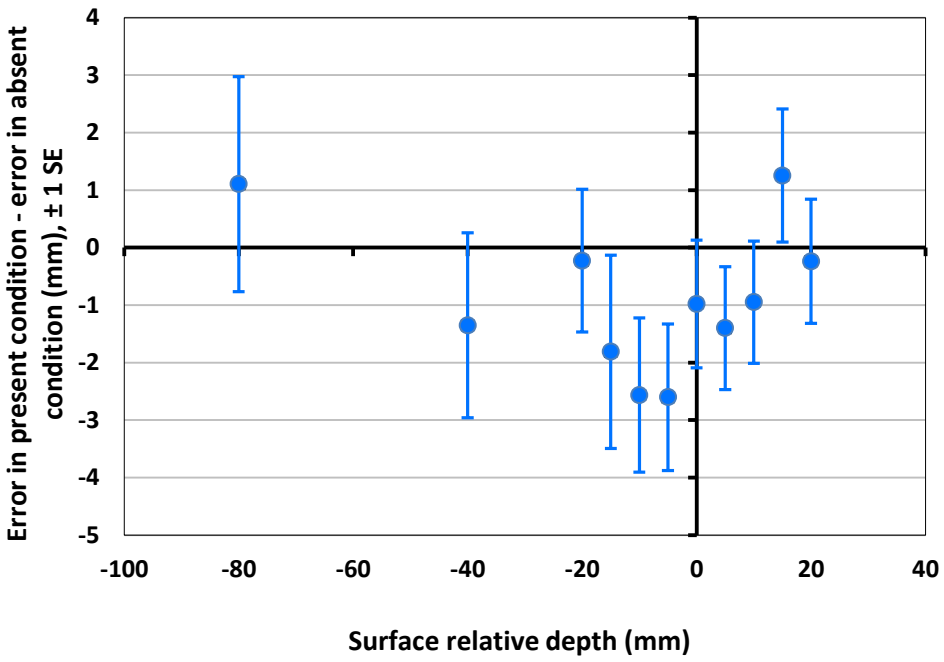

Figure 2.26 The difference between the errors in occluder = present condition and the errors in occluder = absent condition.

The ANOVA analysis showed that the occluder had no main effect, but did have an interaction effect with distance (Figure 2.22). By subtracting the errors in the occluder = absent condition from the errors in the occluder = present condition, we found a push-forward effect around the occluder (Figure 2.26). This shows that the virtual object looks closer to the observer when the occluder is introduced. And, the effect was stronger for the distances behind the occluder than the distances in front of the occluder. This result is consistent with Ellis et al. [1998], where they suggested that the effect was caused by changes in the vergence angle that were influenced by introducing the occluder. When we introduced the occluder, the observer preferred to focus on the occluder, because it was physically present and covered with very salient focal cues. When the virtual object was displayed behind the occluder, the observer's vergence angle increased, and therefore the

virtual object appeared to be closer. For distances of 20mm in front of the occluder, because the virtual object was spinning and its volume was 40mm x 40mm x 40mm, some portion of the virtual object was still behind the occluder, and therefore the occluder still made the virtual object look closer.

Compared with Experiment I, Experiment II's data in Figure 2.22 supports the idea that there was a bias effect in Experiment I. In addition, Experiment II does not support the assumption in Edwards et al. [2004] that there was no average error for distances in front of the occluding surface.

2.4 Experiment III

2.4.1 Design

In Experiment III, we recruited 3 expert subjects who are depth perception researchers—the author of this thesis, her advisor, and one of her committee members. All the subjects were not naive about the purpose of the study, and therefore we can know that the three subjects carefully performed each trial of depth matching. We used the same within-subject design as Experiment II. Each subject performed depth matching for 5 repetition in a row for each distance, with 2 repetitions of 11 distances in the occluder = present condition (5 x 2 x 11=110 trials) and 1 repetition of 11 distances in the occluder = absent condition (5 x 1 x 11=55 trials). We chose fewer trials for the occluder = absent condition, because Experiment II showed that the judged distances were much more stable in the absent condition than in the present condition. Considering the fatigue of subjects performing hundreds of trials, we preferred fewer trials in the absent condition.

Unlike in Experiment II, we measured each subject's IPD at 400mm, and used this value in both the software and the haploscope. Before the experiment, we carefully calibrated the haploscope for each subject using this IPD. Also, we improved the depth indicator. We replaced the green LED with a sharp pointer, which we positioned just underneath the virtual object. Therefore, the 1.5cm gap between the indicator and the virtual object in Experiment II was reduced to just a few millimeters. Another important aspect of the subjects is that they span an interesting range of ages: subject S.E. is 67, subject E.S. is 49, and subject C.H. is 26. Therefore, we assume that the accommodative ability of S.E. is weaker than E.S., and E.S. is weaker than C.H. After finishing the sessions run with S.E., we covered one light above the haploscope system and made the environment dimmer for subjects E.S. and C.H.

2.4.2 Result

During the experiments, we found a warm up process for subjects S.E. and E.S. Within the first 5 trials, these two subjects had errors of 1cm to 2cm, but then they started matching accurately. Figure 2.27 shows the learning process for each subject (not including warm up trials: Figure 2.27 does not show the warm up process because we restarted collecting the data once we found the warm up process). The results show that once the subjects passed the warm up process, there was only a small learning effect for subject E.S. in the occluder = present condition, and S.E. in the occluder = absent condition.

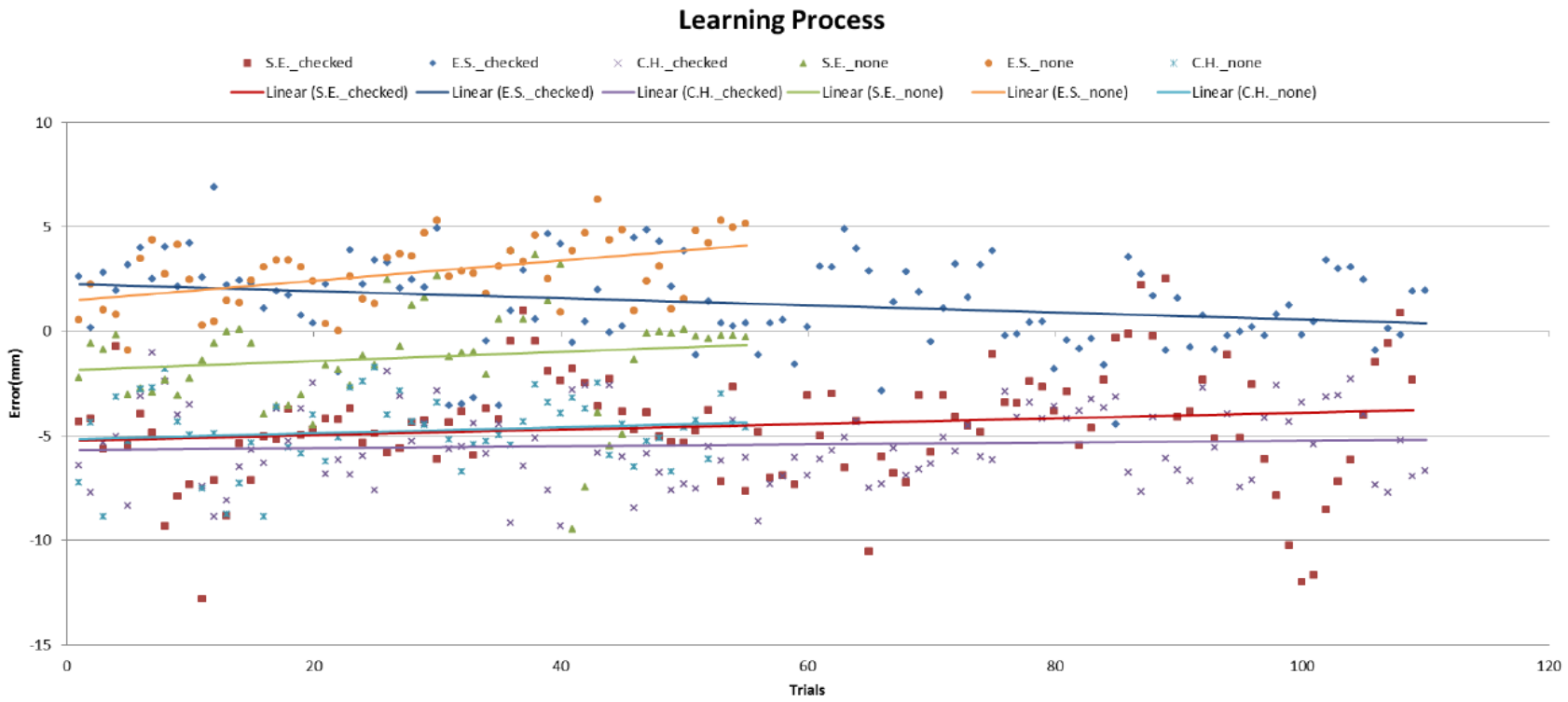

Figure 2.27 Learning process of three subjects

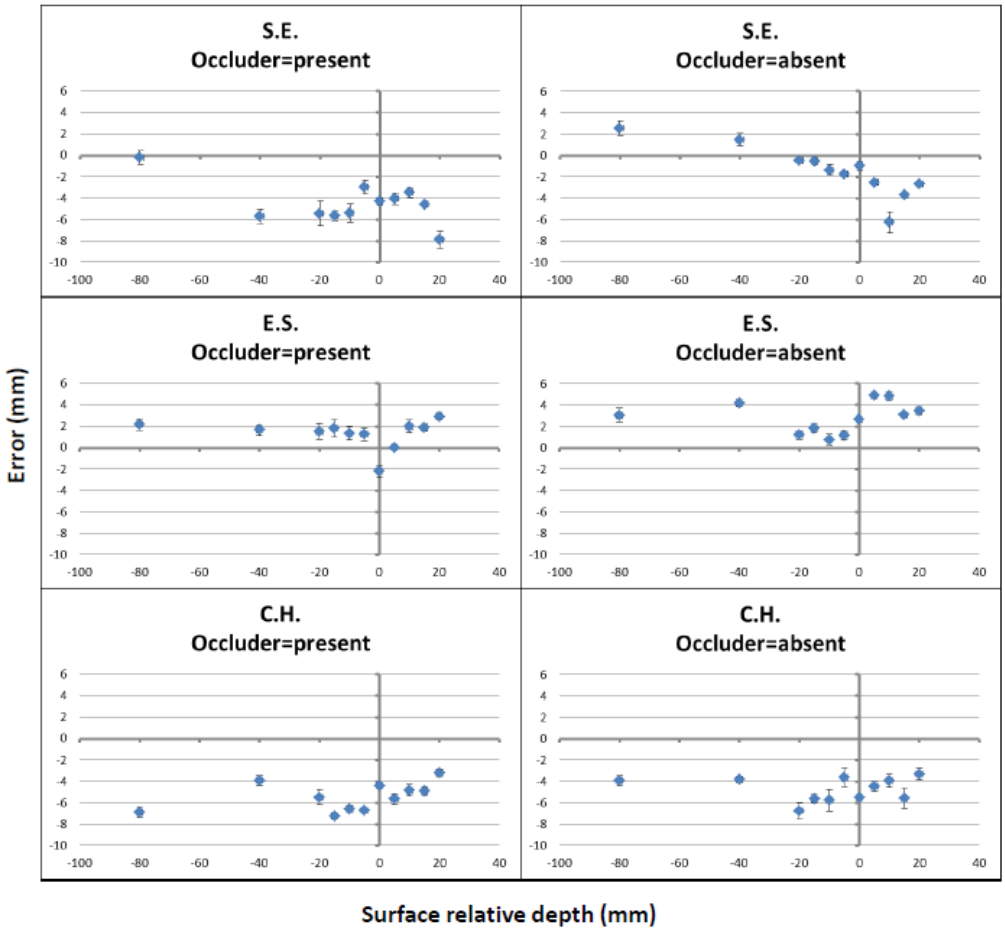

Figure 2.28 Error by observer and occluder condition in Experiment III

Figure 2.28 shows the error of each subject by occluder condition. S.E. underestimated up to 8mm when the occluder was present, and underestimated up to 6mm when the occluder was absent. C.H. also underestimated, with an error up to 8mm in both conditions. Different from the other two subjects, E.S. overestimated in both conditions, with error up to 5mm. And interestingly, he was more accurate in the occluder = present condition than in the occluder = absent condition.

In order to examine the effect of introducing the occluder for each subject, Figure 2.29 shows the differences between mean errors in the two occluder conditions. For S.E., the result was consistent with the results in Experiment II; S.E. observed the virtual object to be closer when the occluder was introduced, and the push-forward effect was larger at distances behind the occluder. E.S. only observed the virtual object closer when it was displayed in front of the occluder, right at the occluder, and also at depths of –40mm and –80mm. C.H. observed the virtual object closer when it was displayed at the depths of –80mm, –15mm, –10mm, –5mm, 5mm, and 10mm, but not at the depths of –40mm, –20mm, 0mm, 15mm, and 20mm.

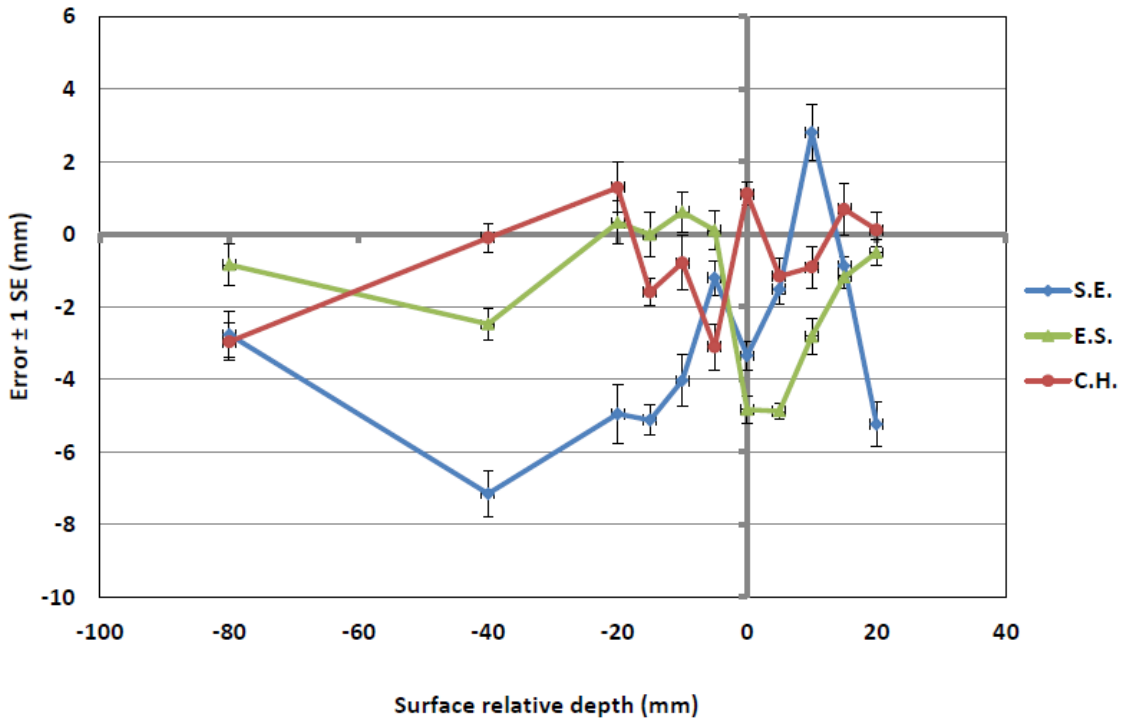

Figure 2.29 he difference in error in the present condition and the error in the absent condition by subject in Experiment III

2.4.3 Discussion

Comparing with the errors in Experiment II, the range of the errors of the three subjects (–8mm ~ 5mm) in this experiment was consistent with that of Experiment II (–6mm ~ 2mm). Therefore, Experiment III confirms that the calibration in Experiment II was accurate.

An interesting phenomenon is that the result from S.E. is very similar to what we found in Experiment II: underestimation in both occluder conditions, and the occluder had a push-forward effect on the virtual object. Also, in the occluder = absent condition, S.E.'s underestimation decreased as the virtual object's distance increased, again consistent with Experiment II. S.E. was the oldest subject.

Subject E.S. overestimated the depth of the virtual object on average, which was the opposite of the underestimation found in Experiment II. Also, for E.S. the occluder did not have a significant effect for distances up to 20mm behind the occluder, but did show a push-forward effect for all 5 distances in front of and right on the occluder.

For subject C.H., the effect of the occluder was noisy, sometimes showing a push-forward effect, but sometimes showing push-back effect. C.H. was the youngest subject. Although the age of C.H. was similar to the average age of subjects in Experiment II, her results are not consistent with Experiment II.

To better understand the reasons behind these phenomena, one idea is to examine the subjects' vergence when they are observing the virtual object, in both occluder conditions. If the vergence change is consistent with the change in errors after the introduction of the occluder, then vergence change is a critical factor behind the results.

Because the subjects spanned a range of ages, we suspected that the subjects' differing accommodative ability would have an effect in the results. However, our results were not consistent with this hypothesis. Because the accommodative demand was always set to 400mm, we would expect underestimation for distances behind the occluder. Also, considering the push-forward effect due to convergence change, C.H. should have the strongest underestimation effect for depth behind the occluder. However, the results show that the strongest underestimation effect happened for subject S.E., who is the oldest subject. Therefore, we doubt that accommodation ability had a significant effect in this experiment.

It is also important to mention that we made the environment dimmer for subject E.S. and C.H. compared with S.E., which might result in a brighter virtual object. According to Singh [2013], brighter virtual objects appear closer. However, the result showed that the virtual object looked farther away for E.S. compared with S.E. The reason might be that the dimmer environment did not result in a brighter virtual object.

CHAPTER III

CONCLUSION AND FUTURE WORK

This work was intended to study the effect of an occluding surface on the accuracy of near field depth perception in augmented reality. Three experiments were designed and conducted on our haploscope, which can independently vary accommodation and vergence. Experiment I applied a between-subject design, and found neither a main effect of occluder, nor an occluder by distance interaction effect, on the accuracy of depth matching. Comparing the results to Edwards et al. [2004], our results were more accurate after we applied Edwards et al.'s normalization method. However, our results also suggested an error bias for our data, which was likely caused by insufficient calibration.

Therefore, we conducted Experiment II. Here, we improved the calibration method, and we used a within-subject design. The results still showed an interaction of occluder by distance on the accuracy of depth matching. In addition, subjects tended to observe the virtual target closer after the occluder was introduced, both for distances in front of the occluder and behind the occluder. This result was consistent with Ellis et al. [1998], and suggests that the occluder had an effect on subject's vergence. In addition, we found underestimation for both conditions at all distances but –80mm. A likely reason for this is that we used the subjects' IPD measured at optical infinity, but IPD decreases as observers look at nearby objects. This suggests a future experiment, which basically

replicates Experiment II, but dynamically calculates the observer's IPD based on the distance to the virtual target.

For Experiment III, all three subjects were depth perception researchers, and we used their IPD at 400mm. Only subject S.E.'s result was consistent with that of Experiment II. Our results were not explainable by hypothesis based on the ages of the subjects, nor based on the luminance of the environment.

In the future, this thesis suggests conducting experiments that study the effect of the near IPD change, the effect of age, and the effect of environmental luminance. In addition, our current haploscope is physically attached on a table, and therefore only supplies one rigid view direction. In many real AR-assisted medical applications, wearable AR equipment would likely be used. Therefore, it would be interesting to make a wearable haploscope—basically, an HMD with adjustable focal and vergence demand—to further study the effect of an occluding surface on depth matching.

REFERENCES

Abdel-Aziz, Y. I. and Karara, H. M. [1971]. Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-range Photogrammetry. *Proceedings: The symposium on Close-Range Photogrammetry* (pp. 1-18). Falls Church, VA: American Society of Photogrammetry.

ART, [2012]. System User Manual SmartTrack & DTrack. Version 2.7.

ARToolKit Documentation, Retrieved. [March 2014], HIT Lab, University of Washington, http://www.hitl.washington.edu/artoolkit/.

Axholt, M. [2011]. Pinhole Camera Calibration in the Presence of Human Noise. Ph.D. thesis, Linköping University, Norrköping, Sweden.

Azuma, R. [1997]. A Survey of Augmented Reality. *Teleoperators and Virtual Environments* (vol. 6, no. 4, pp. 757-764).

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S. and MacIntyre, B. [2001]. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, (vol. 21, no. 6, pp. 34-47).

Bajura, M., Neumann, U. [1995]. Dynamic Registration Correction in Video-Based Augmented Reality Systems. *IEEE Computer Graphics and Applications* (pp 52-60).

Bichlmeier, C., Ockert, B., Heining, S. M., Ahmadi, A., Narvab, N. [2008]. Stepping into the Operating Theater: ARAV – Augmented Reality Aided Vertebroplasty. *ISMAR* (pp. 165-166).

Bingham, G. P., Bradley, A., Bailey, M., and Vinner, R. [2001]. Accommodation, Occlusion,and Disparity Matching Are Used to Guide Reaching: A Comparison of Actual Versus Virtual Environments. *Journal of Experimental Psychology: Human Perception and Performance* (vol. 27, no. 6, pp. 1314–1334).

Bridgeman, B., Peery, S., and Anand, S. [1997]. Interaction of Cognitive and Sensorimotor Maps of Visual Space. Perception & Psychophysics, (vol. 59, no. 3, pp.456–469).

Ciuffreda, K. J. [1992]. Components of Clinical Near Vergence Testing. Journal of Behavioral Optometry (vol. 3, pp. 3-13).

Creem-Regehr, S. H., Willemsen, P., Gooch, A. A., and Thompson, W. B. [2005]. The Influence of Restricted Viewing Conditions on Egocentric Distance Perception: Implications for Real and Virtual Indoor Environments. Perception (vol. 34, pp.191–204).

Cutting, J. [1997]. How the eye measures reality and virtual reality. Behavior Research Methods (vol. 29, pp. 27–36).

Cutting, J. E. and Vishton, P. M. [1995]. Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information About Depth. In W. Epstein and S. Rogers (Eds), Handbook of Perception and Cognition: Perception of Space and Motion (vol. 5, pp. 71–117). New York: Academic.

Dou, M., Fuchs, H. [2014]. Temporally Enhanced 3D Capture of Room-sized Dynamic Scenes with Commodity Depth Cameras. IEEE Virtual Reality.

Drascic, D. and Milgram, P. [1996]. Perceptual Issues in Augmented Reality. Stereoscopic Displays and Virtual Reality Systems III (vol. 2653, pp. 123–134).

Edwards, P. J., Hawkes, D. J., Hill, D. L., Jewell, D., Spink, R., Strong, A., Gleeson, M. [1995]. Augmentation of Reality in the Stereo Operating Microscope for Otolaryngology and Neurosurgical Guidance. Journal of Image-Guided Surgery (vol. 1, no. 3, pp. 172-178).

Edwards, P. J., Johnson, L. G., Hawkes, D. J., Fenlon, M. R., Strong, A. J. and Gleeson, M. J. [2004]. Clinical Experience and Perception in Stereo Augmented Reality Surgical Navigation. Medical Imaging and Augmented Reality (pp. 369376).

Edwards, P. J., King, A. P., Hawkes, D. J., Fleig, O., Maurer, C. R., Hill, D. L., Fenlon, M. R., Cunha, D. A. de, Gaston, R. P., Chandra, S., Mannss, J., Strong, A. J., Gleeson, M. J.,. Cox, T. C. [1996]. Stereo Augmented Reality in the Surgical Microscope. Journal of Studies in health technology and informatics (vol. 62, pp. 102-108).

Ellis, S. R. and Menges, B. M. [1998]. Localization of Virtual Objects in the Near Visual Field. Human Factors (vol. 40, no. 3, pp. 415–431).

Fisher S. K., and Ciuffreda K. J. [1988]. Accommodation and Apparent Distance. Perception (vol. 17, pp. 609-621).

Fuchs H., Livingston M. A., Raskar R., Colucci D., Keller K., State A., Crawford J. R., Rademacher P., Drake S. H., Meyer A. A. [1998]. Augmented Reality Visualization for Laparoscopic Surgery. Proceedings: the First International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 934-943).

Hansen, C., Wieferich, J., Ritter, F., Rieder, C., Peitgen, H-O [2010]. Illustrative Visualization of 3D Planning Models for Augmented Reality in Liver Surgery. CARS (vol. 5, no. 2, pp. 133-141).

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., and Fitzgibbon, A. [2011]. KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. Proceeding: ACM Symposium on User Interface Software and Technology.

Johnson, L., Eadie, L., Cunha, D. de, Edwards, P., Hawkes, D. [2001]. The Limitation of Transparent Depth Perception with Restricted Field of View: Application in an Augmented Reality Surgical Microscope. Proceedings: the Human Factors and Ergonomics Society Annual Meeting (vol. 45, no. 8, pp. 1438-1442).

Kelly, P. J., Kall, B. A., Goerss, S., Earnest, F. [1986]. Computer-assisted Stereotaxic Laser Resection of Intra-axial Brain Neoplasms. Journal of Neurosurgery (vol. 64, pp. 427-439).

Kersten-Oertel, M., Jannin, P., Collins, D. L. [2013]. The State of the Art of Visualization in Mixed Reality Image Guided Surgery. Computerized Medical Imaging and Graphics (vol. 37, pp, 98-112).

Krishnan, V. V., Phillips, S., Stark, L. [1973]. Frequency Analysis of Accommodation, Accommodative Vergence, and Disparity Vergence. Vision Research (vol. 13, pp. 1545-1554).

Lepetit, V., Vacchetti, L., Thalmann, D., Fua, P. [2003]. Fully Automated and Stable Registration for Augmented Reality Applications. ISMAR (pp. 93-102).

Loomis, J. M. and Knapp, J. M. [2003]. Visual Perception of Egocentric Distance in Real and Virtual Environments. In L. J. Hettinger and M. Haas (Eds), Virtual and Adaptive Environments: Applications, Implications,and Human Performance (chapter 2, pp. 21–46). CRC Press, Mahwah, NJ.

Mccandless, J.W., Ellis, S.R., and Adelstein, B.D [2000]. Localization of a Time-Delayed, Monocular Virtual Object Superimposed on a Real Environment. Presence (Camb) (vol. 9, no. 1, pp. 15–24).

Moons, T., Gool, L. V., Vergauwen, M. [2010]. 3D Reconstruction from Multiple Images Part 1: Principle. Journal of Foundations and Trends in Computer Graphics and Vision (vol. 4, no. 4, pp. 287-404).

Scheuering, M., Schenk, A., Schneider, A., Preim, B., Greiner, G. [2003]. Intraoperative Augmented Reality for Minimally Invasive Liver Interventions. SPIE Medical Imaging (vol. 5029, pp. 407-417).

Semlow, J. L., Hung, G. K. [1983]. The Near Response: Theories of Control. In C. M. Schor & K. J. Ciuffreda (Eds), Vergence Eye Movements: Basic and Clinic Aspects (pp. 175-196).

Sielhorst, T., Feuerstein, M., and Navab, N. [2008]. Advanced Medical Displays: A Literature Review of Augmented Reality. Journal of Display Technology (vol. 4, no. 4, pp. 451–467).

Singh, G. [2013]. Near Field Depth Perception in Optical See-though Augmented Reality. Ph.D. Dissertation, Dept. of Computer Science and Engineering, Mississippi State University.

Suthau, T., Vetter, M., Hassenpflug, P., Meinzer, H. P., Hellwich, O. [2002]. A Concept Work for Augmented Reality Visualization Based on a Medical Application in Liver Surgery. ISPRS Commission V Symposium.

Sutherland, I. [1968]. A Head-Mounted Three-Dimensional Display. Proceeding: AFIPS, Fall Joint Computer Conference (pp. 757-764), Thompson Books, Washington, D.C.

Suzuki, N., Hattori, A., Hashizume, M. [2008]. Benefits of Augmented Reality Function for Laparoscopic and Endoscopic Surgical Robot Systems. AMI-ARAC.

Tobias, S., Brief, J., Welzel, T., Giesler, B., Hassfeld, S., Muehling, J., Dillmann, R. [2002]. INPRES (Intraoperative Presentation of Surgical Planning and Simulation Results) – Augmented Reality for Craniofacial Surgery. Surgetica.

Total Immersion, Retrieved, [2014], http://www.t-immersion.com/about-us

Tsai, R. Y. [1986]. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. Proceedings: IEEE Conference on Computer Vision and Pattern Recognition (pp. 364-374), Miami Beach, FL.

Virtual Dressing Room, TryLive solution, Retrieved, [2014], Total Immersion, http://www.trylive.com/solutions/trylive-apparel-virtual-dressing-room.

Vuforia product, Retrieved, [2014], Qualcomm Connected Experiences, Inc, https://www.vuforia.com/.

Zhang, Z. [2000]. A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence (vol. 22, no. 11, pp. 1330-1334).