# Tablet versus Phone: Depth Perception in Handheld Augmented Reality

[1]Arindam Dey* [1]Graeme Jarvis† [1]Christian Sandor‡ [2]Gerhard Reitmayr§

[1]Magic Vision Laboratory, University of South Australia
[2]Institute of Computer Graphics and Vision, Graz University of Technology

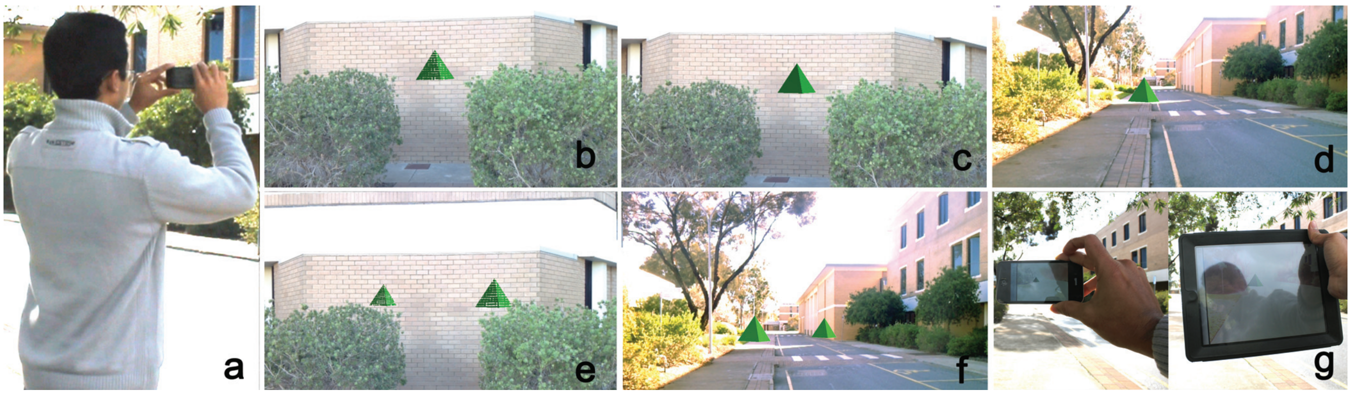



Figure 1: Experimental setup (a): Participants held the devices at their eye height and generally within a comfortable distance (30 cm to 45 cm). In Experiment 2, participants had to estimate the distance to a virtual pyramid in three different conditions. In the first condition the virtual pyramid was shown as occluded by the wall in front and an edge-overlay X-ray visualization was rendered to show the virtual pyramid (b), in the second condition the virtual pyramid was occluded by the wall but we did not render any X-ray visualization and the pyramid appeared to be floating on the wall (c), and in the third condition the virtual pyramid was shown without any occlusion (d). For Experiment 3, only (b) and (d) were used, but we varied the display devices additionally. In Experiment 4 we evaluated exocentric and ordinal depth perception. As a stimulus, we showed two pyramids with (e) and without occlusion (f). All of our experiments were conducted using an iPad and an iPhone (g).

## Abstract

Augmented Reality (AR) applications on mobile devices like smartphones and tablet computers have become increasingly popular. In this paper, for the first time in the AR domain, we present: (1) the influence of different handheld displays and (2) the exocentric depth perception. Unlike egocentric depth perception, exocentric depth perception has not been investigated in AR.

We have selected a suitable vision-based tracking method for our user studies based on a set of evaluations. Then we have investigated the effect of display size and resolution through two user studies. One study investigated the effect of different displays on egocentric depth perception. The other study investigated the effect of displays on exocentric and ordinal depth perception.

Interestingly, we noticed depth compression is less when using a mobile phone, while participants subjectively preferred a tablet. A similar effect was also noticed in exocentric depth perception. The tablet provided significantly better ordinal depth perception and faster response time than the mobile phone. In both of the studies, we found no effect of the AR X-ray visualization on depth perception. Both egocentric and exocentric distances were underestimated.

**Keywords:** Augmented Reality, User Evaluation, X-ray Visualization, Handheld Displays, Outdoor Environment, Depth Perception

**Index Terms:** H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities H.1.2 [Models and Principles]: User/Machine Systems—Human factors

*e-mail: arindam.dey@unisa.edu.au
†e-mail: fanged@gmail.com
‡e-mail: christian@sandor.com
§e-mail: reitmayr@icg.tugraz.at

IEEE International Symposium on Mixed and Augmented Reality 2012
Science and Technology Proceedings
5 - 8 November 2012, Atlanta, Georgia
978-1-4673-4661-0/12/$31.00 ©2012 IEEE

## 1 Introduction

AR superimposes computer generated virtual objects on top of the real world with 3D registration, and provides interaction in real time [4]. Typical AR applications are comprised of three fundamental components—display device, tracking, and rendering techniques.

A display presents the AR environment to the user. Hence, it is important to have a high quality display to provide effective visual stimuli. While Head-Worn Displays (HWDs) provide better immersion, portable handheld displays enable higher mobility and make AR widely usable in outdoor environments. Particularly, with the increase in computational power of current portable devices like tablets and mobile phones, it has become a widespread practice to build AR application on these platforms. The introduction of a number of AR Browsers such as [38, 31] and AR development APIs for smartphones in the past few years support this claim.

A precise tracking of the camera within the augmented environment is required to achieve proper alignment of the virtual objects to the real-world counterparts and create a rich user experience [2]. A broad range of research is being carried out to develop highly accurate and robust tracking technologies. Sensor-based tracking is widely used to track the pose of a mobile phone's camera in an outdoor location. In contrast to the sensor-based tracking for outdoor AR, recently, Kim et al. [17] have proposed a tracking method based on a set of keyframes that are selected from a video stream for mobile AR. As achieving a proper registration is one of the fundamental challenges in mobile AR [5], we believe this challenge has prevented scientists from conducting reliable user studies on handheld AR in outdoor locations.

Rendering techniques are used to present virtual objects to the user in an AR environment with a proper context to the real world. One such rendering technique is X-ray visualization. X-ray visualizations are very specific to AR and are useful in various application domains, including medical, defense, and tourism. We have previously presented two different kinds of photorealistic X-ray visualizations—edge-based [3] and saliency-based [36]—enabling users to see the photorealistic (in contrast to symbolic) appearance of the background points of interest (POIs), while providing a context of the real-world foreground.

Current research shows a clear lack of user-based research, investigating perceptual capabilities for mobile AR using handheld devices. Most of the previous user studies in AR were conducted in indoor locations, primarily using HWDs. Only a few papers have presented user studies conducted in an outdoor location using a handheld display [7, 40, 8]. However, with the current trend in AR, it has become an absolute necessity to perform more user-based research on outdoor handheld AR. In this paper we have thoroughly investigated the effect of display size and resolution on outdoor depth perception. While egocentric depth perception is widely investigated in AR, exocentric depth perception is overlooked. Additionally, we have investigated the effect of AR X-ray visualization on depth perception.

In Experiment 1, to select a suitable tracking method for our focused user studies, we have conducted a preliminary benchmark to compare the performance of sensor-based and vision-based tracking methods in an outdoor location.

Then, we conducted a user study (Experiment 2) to evaluate the effect of distance-field and AR X-ray rendering on egocentric depth perception using an iPhone. Previously we have noticed a contradictory result on outdoor depth perception; distance was one of the possible factors influencing the difference (see [7]). We also investigated if tracking methods have any influence on depth perception.

Neither Experiment 1 nor Experiment 2 revealed any significant differences between the tracking methods; participants subjectively preferred sensor-based tracking and hence, we have used this method for our two other user studies, investigating the effect of display properties on depth perception.

In Experiment 3, we have systematically varied display size and resolution to investigate their effect on *egocentric* depth perception.

Finally, in Experiment 4, we investigated the effect of display configuration on *exocentric* and *ordinal* depth perception.

We noticed a consistent underestimation of both egocentric and exocentric distances. Surprisingly, the iPhone led to less depth compression in comparison with the iPad. The iPad led to significantly more correct responses in term of ordinal depth perception. While display size indicated an effect on depth perception, we did not notice any effect of display resolution.

### 1.1 Contribution

In the AR domain, we are the first to investigate: (1) the influence of *handheld displays* and (2) the *exocentric* depth perception. While most of the previous depth perception experiments in AR were conducted using HWDs, this is the first experiment to investigate depth perception using a mobile phone and a tablet computer. No other user studies has investigated exocentric depth perception in AR so far. Both issues are very important in the current state-of-the-art in this domain. Handheld devices are the most promising AR platform for mass adoption, and properly perceiving the relationship between two virtual objects is crucial in outdoor AR. We believe that this research will provide important insight to mobile AR application developers and researchers.

The rest of the paper is organized as follows. We review the earlier research conducted in this domain in Section 2. In Section 3, we present the results of a benchmark showing the difference in tracking quality of sensor-based and vision-based tracking methods. Section 4 reports a user study investigating the effect of tracking quality on outdoor depth perception. Next, we present another user study describing the effects of display size and resolution on egocentric depth perception in Section 5. Section 6 reports the effect of display on exocentric and ordinal depth perception. We conclude our paper by directing towards future research in Section 7.

## 2 Related Work

AR, being a user interface technology, has to be evaluated extensively with human participants to understand the underlying perceptual and cognitive properties. There have been a number of studies evaluating depth perception in AR. In this section we review some of the relevant work related to our current experiments and experimental setups including tracking technology for outdoor AR.

### 2.1 Depth Perception and Depth Cues

Perception of distance helps us to create a three-dimensional impression of the world by combining two two-dimensional, flat retinal images captured from slightly different viewpoints [39]. The three dimensions of the perceptual space are *up-down*, *sideways*, and *towards-away*.The towards-away dimension presents the depth information to our brain and we receive less information in this dimension in comparison with the other two dimensions [43]. Our brain has to indirectly infer the information—*depth cues*—received in that dimension.

There are two categories of depth cues—pictorial and non-pictorial. In the real world, depth perception is based on the interpretation of one or more depth cues. Ten important depth cues in particular influence us [14].

These cues change their strength based on distances [6, 32]. Our environment can be divided in three different distances as near-field (within arm's reach), medium-field (up to 30m), and far-field (beyond 30m) [6]. The interaction and combination of these depth cues are described in *cue theory* [20].

### 2.2 Depth Perception in AR

Understanding depth perception in AR is necessary to portray the correct relationship between the real-world objects and the virtual objects to the observer as intended by the developer. A wide range of research has been carried out on this domain.

Perception is an invisible cognitive state and we measure the perception of depth through quantifiable depth judgments and perception is inferred from these judgments [39]. To measure the depth judgment scientists have employed different types of tasks including verbal estimation, closed- and open-loop action based tasks. A review of such tasks are presented by Loomis and Knapp [25].

Two different types of depth judgment occur in our environment—*egocentric* and *exocentric* [30]. Egocentric depth perception refers to the distance to an object perceived from the observers viewpoint; exocentric depth perception refers to the distance between two objects in the view [39]. Both egocentric and exocentric depths are underestimated in virtual environments [33]; while in the real world, depth estimation is somewhat accurate [25]. To solve this problem various synthetic depth cues have been proposed [24, 40, 44].

**Egocentric Depth Perception**

So far AR has been investigated widely, specifically using HWDs. Most of the depth perception studies in AR, like Virtual Reality, have consistently reported depth underestimation of objects presented on a ground plane. However, the reason of this underestimation is not clearly understood.

Numerous studies have evaluated egocentric depth perception in near-field distances. The effect of near-field distances along with an occluded surface, convergence, accommodation, age, and stereo displays were studied through a perceptual matching task in a series of studies [10]. Later on, the effect of motion parallax and system latency was explored in [28]. Recently, a study experimented reaching and matching tasks in near-field distances [37].

Egocentric depth judgment in medium- and far-field AR was evaluated using a perceptual matching protocol in [39]. This experiment interestingly reported a shift in bias from underestimation to overestimation at 23m in an indoor environment, whereas depth underestimation is a common phenomenon in virtual environments. Later on, another experiment without using any X-ray vision, reported depth overestimation of medium-field distances in an outdoor environment [22]. Multiple visualizations for occluded objects were evaluated in [24, 23]. Recently, the effect of peripheral vision was evaluated [15]. All of the above studies used an optical see-through HWD.

However, in the last few years, handheld devices like mobile phones, have become a promising platform for AR applications, as their computational power increased [19]. While perceptual issues in AR are investigated predominantly using HWDs, handheld displays remained under-explored.

Some recent studies, investigated perceptual issues with X-ray visualization using handheld displays. A set of depth cues for

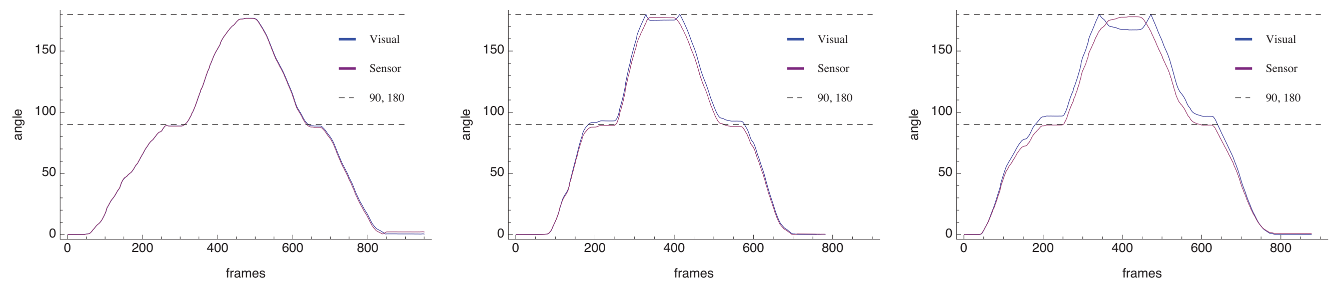



Figure 2: Angular measurements from the visual tracker and sensors for different distances from the axis of rotation. Left 0% offset, middle 5% and right 10%. Beyond 180°, the rotation is measured in the other direction and the angle decreases.

X-ray visualization was evaluated in [40]. We have investigated X-ray visualization with a target selection task in [36], with a depth perception task in [7], and with a real-world navigation task in [8]. Results in [7], contradicting [22], reported a consistent depth underestimation in an outdoor environment, however, of far-field distances. In this paper, we are further investigating if different range of distances affect depth perception.

**Exocentric Depth Perception**

Numerous experiments have investigated the exocentric depth perception in real-world scenario. It was reported that the change in the viewing angle can change the exocentric depth perception of exactly same stimuli [21]. A mathematical model of the visual space was created by Foley et al. based on exocentric depth judgment [12]. Loomis et al. proved a dissociation between perceived target location and perceived exocentric distance and shape [26]. The effect of familiar and unfamiliar size of objects were investigated by Predebon [34].

While the AR community has conducted egocentric depth perception experiments widely, exocentric depth perception still needs to be explored in this space. Exocentric depth perception is important for AR applications where relationship between two or more virtual objects are to inferred.

Unavailability of robust tracking methods for outdoor AR makes it harder for researchers to conduct reliable user studies in such environments.

### 2.3 Tracking Methods in AR

Estimating and continuously tracking the user's or camera's pose is an essential pre-requisite for any augmented reality application. For location-based AR Browsers, sensor-based approaches combining GPS for position and magnetometer and IMUs for orientation have been used for a long time [11]. With modern mobile devices featuring the same set of sensors, this is the most common approach to the tracking problem and widely used in mobile applications [38, 31]. Computer vision-based approaches are mature for tracking objects relative to the user's devices, ranging from fiducial markers [16] to arbitrary image targets [42].

For outdoor environments, computer vision-based tracking systems are feasible, but require larger models [35] and often combine sensor input for more robust performance [45]. The largest source of overlay error is usually inaccurate orientation estimation and for a stationary user looking around this is the most important concern. Stable and fast orientation estimation from a rotating camera without prior information about the environment is possible [9, 27] and efficient variations exist for mobile devices [41, 17]. Without further information, the visual orientation tracking is only relative. This is offset through combination with absolute sensors such as magnetometers and linear accelerometers.

Overall, even though mobile devices are a highly sought after platform for commercial AR applications, the current state of the literature indicates a clear lack of experiments that investigate the effects of display on depth perception in mobile AR. We did not find any user study investigating exocentric depth perception in AR. To fill this gap, in this paper, we have conducted multiple evaluations investigating egocentric and exocentric depth perception using handheld displays.

## 3 Experiment 1: Difference in Tracking Quality

We conducted a preliminary benchmark to compare the performance of sensor-based and vision-based tracking methods in an outdoor location to select a suitable tracking method for our focused user studies.

In our system we use the visual orientation tracker described by Kim et al. [17]. This tracker computes the device orientation with respect to a set of keyframes that are selected from the video stream and added over time similar to well-known systems, such as PTAM [18]. The initial keyframe determines the reference orientation. Each incoming video frame is tracked by re-projecting visible interest points from near-by keyframes into the current frame. Then active search in the neighborhood of the re-projection locations is used to establish correspondences. The camera orientation is estimated by minimizing the re-projection error over all valid correspondences. As the motion model only allows for camera rotation, the visual tracker does not need to estimate the 3D locations of the interest points.

Such tracking approaches have been demonstrated to work efficiently and accurately even on mobile platforms [41]. In order to assess the need for such methods, we ran an evaluation to compare the accuracy of estimated orientations between the visual tracker and built-in orientation sensors. Specifically, we used the CoreMotion API on an iPhone4S device to extract orientation measurements and use these independently from the visual orientation estimation.

### 3.1 Setup

The evaluation setup consist of a turn-table with the phone mounted vertically in landscape mode such that the camera center is in the axis of rotation. Around the turn-table we set up a scene structure at 1m distance. The phone was mounted such that it could be translated radially outwards along its viewing direction to emulate the effects of off-axis rotation. We verified that the turn-table did not influence the magnetometer measurements and the resulting orientation estimates. Repeated measurements of the same directions as well as relative rotations were stable.

We varied the distance of the camera focal point from the axis of rotations to simulate the effects of a user holding the phone and rotating around their body axis. The phone was moved to 0cm, 5cm and 10cm from the axis corresponding to a relative radius of 0%, 5% and 10% of mean scene distance. For a real user, holding the phone around 50cm from the axis of rotation this corresponds to a scene distance of ∞, 10m, and 5m. In each setting we rotated the phone through a rotation of 180°and measured the relative rotation from the starting pose using both the visual tracker and the build-in sensors (IMUs). We recorded 3 runs for each setting.

### 3.2 Results

We looked at both static errors for 2 rotations at 90°and 180°and dynamic errors between the two measurement methods. We compared the angle of rotation to account for any inaccuracies in aligning the sensor or camera coordinate system with the turn-table axis of rotation. Figure 2 shows the rotation angles for three recordings, each for a different camera distance from the axis of rotation.

Table 1 gives the mean and standard deviation for the different configurations and settings. For the perfect motion at 0% distance from the rotation axis, the accuracy for both the visual tracking

Table 1: Angular measurement accuracy comparing the visual tracker and sensor-based orientation estimates. The visual tracker shows a clear trend with increasing off-axis motion, while the sensor-based orientation estimate is stable.

| Offset | 90° | | 180° | |
|---|---|---|---|---|
| | Visual M(SD) | Sensor M(SD) | Visual M(SD) | Sensor M(SD) |
| 0% | -1.20(0.11) | -1.40(0.12) | -2.92(0.19) | -3.36(0.15) |
| 5% | 2.82(0.12) | -0.73(0.17) | 4.77(0.10) | -2.55(0.20) |
| 10% | 6.87(0.12) | -0.63(0.14) | 12.64(0.16) | -2.56(0.19) |

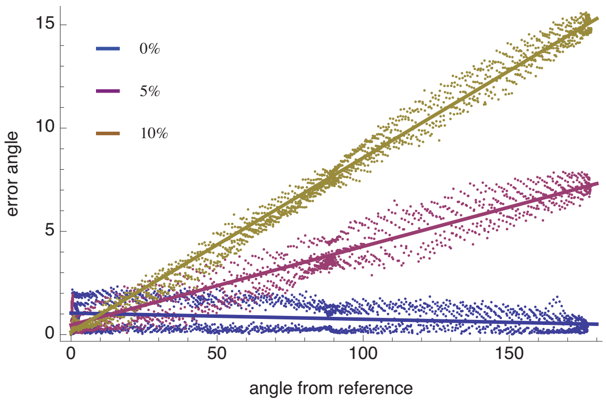



Figure 3: Correlation between the error of the visual tracker and the angle to the reference frame.

and the sensor-based rotation is to 1.5° and about 3° for 90° and 180° rotations. Both exhibit a slight underestimation of the orientation. For offsets from the rotation axis, the visual tracking shows a marked drift which is due to the translation in the image being modeled as additional rotation and therefore the orientation is overestimated consistently.

We can see that for pure rotation, or rotations close to the initial reference direction, both methods perform very similar. Therefore, we do not expect to see any influence on the user's performance. Moreover, the sensor-based method also provides absolute orientation which can be more useful than a pure relative orientation.

Deviations from the pure rotational model are visible and are due to the parallax effect induced by the camera motion. From the given data, we estimated the drift rate for the visual tracker by correlating the angular difference between the two rotations and the amount of total rotation as estimated by the sensors (see Figure 3). For a 5% offset, we see a rate of 0.038, while for the 10% offset a rate of 0.084, roughly twice as much. This implies that the error of the visual tracker is roughly 4% and 9% of the angle to the initial reference for these offsets.

## 4 Experiment 2: Effect of AR X-ray and Distance-field on Depth Perception

In this experiment, we investigated the effect of X-ray visualization and distance-field on egocetric depth perception using an iPhone. Experimentation of depth perception has not widely been performed using handheld displays. Additionally, we have investigated the effect of tracking methods on depth perception, although we did not expect any significant difference based on the results of Experiment 1.

### 4.1 Experimental Platform

We used an iPhone4S with 3.5" diagonal screen having a resolution of $960 \times 640$ as the experimental platform. For more detailed specifications please refer to [1]. While most of the other depth perception studies in AR have used HWDs or other handheld displays with larger screen size, we decided to use a mobile phone as this is going to be the most commonly used platform for outdoor AR applications. User responses were recorded by an experimenter on a laptop placed next to the participant.

### 4.2 Participants and Procedure

27 students and staff, ages ranging from 21 to 63 years, from the University of South Australia were recruited for the *mixed-factorial* experiment and were equally distributed into three matched groups. Each of the groups performed their task on the assigned *Spatial arrangement*, a between subject variable. All the participants had the normal or corrected to normal vision. 18 of them had experience with AR environments.

Participants were instructed to stand at a fixed position during the experimental sessions and hold the phone up at eye height as shown in Figure 1a. We instructed participants to use the display as they would do normally. We noticed that participants held the device within a comfortable range of 30 cm to 45 cm from their eyes. The experimental trials were sequentially presented on the screen of the mobile phone and only one green pyramid was shown at a time. Participants had to verbally report the egocentric distance to the tip of the pyramid. They were aware of the actual size of the pyramid being $3m \times 3m$ base and $3m$ high.

Except occlusion and height in the visual field, no other depth cues (such as shadow or texture gradient) were available to the participants. As we only used an orientation tracker, motion parallax was not available as a depth cue, either.

After reporting the distance to the experimenter, participants tapped twice on the screen and the next trial was presented. This process was repeated 12 times in a single repetition and the distance to the target pyramid was changed randomly. Overall, there were four repetition per participant. Participants were allowed to take a break between each repetitions.

The experiment took about 25 minutes per participant and the whole experiment was conducted over five consecutive days at our university campus.

### 4.3 Variables

This $mixed-factorial$ experiment was based on four independent and four dependent variables. Overall, there were 3 (spatial arrangement) $\times$ 2 (tracking method) $\times$ 12 (distance) $\times$ 2 (repetitions for each tracking method) $\times$ 9 (participants per group) = 1296 data points.

**Independent Variables**

- **Spatial Arrangement** (Combination of Occlusion and X-ray Visualization) $\in$ {On-On, On-Off, Off-Off} *between subjects*

  We presented the target green pyramid as a part of the background scene for two of the three spatial arrangements. Hence, these conditions have the occlusion *On* (Figure 1b, 1c).

  **On-On:** In one of the *occlusion-on* conditions, we have used an edge-overlaid X-ray visualization that preserves the edges of the foreground and shows the background through that edge-overlaid representation of the foreground [3]. This combination constructed one level of this independent variable—*On-On* (See Figure 1b).

  **On-Off:** In another condition with *occlusion-on*, we did not present any X-ray visualization and occluded target object appeared to be floating over the foreground. This combination constructed another level—*On-Off* (See Figure 1c). This type of setup is widely used in various AR browsers.

  **Off-Off:** As a control condition, we presented the target object without any occlusion(*occlusion-off*); and consequently, no X-ray visualization was applied. This combination constructed the final level—*Off-Off* (See Figure 1d).

- **Distance** $\in$ {19.3m (1.86°), 24.2m (1.48°), 28.9m (1.2°), 33.8m (1.04°), 38.6m (0.92°), 47.3m (0.74°), 59.1m (0.62°), 69.7m (0.5°), 82.5m (0.42°), 95.1m (0.36°), 104.6m (0.34°), 117.0m (0.26°)} *within subjects*

We have selected 12 different distances that cover a large range. Distances up to 38.6 meters were selected from an experiment by Livingston et al. [22] where they reported an *overestimation* of distances in an outdoor environment. Distances from 69.7 meters were selected from our previous experiment where, contradicting [22], we found an *underestimation* of distance in an outdoor environment [7]. Distances 47.3 and 59.1 were selected as bridging factors between the two different sets of distances.

The visual size of the target pyramids—the angular size the pyramids take on the retina—at various distances is given within the parentheses (assuming that participants held the device at 45 cm away from their eyes). If the pyramids were present in the real world, then their actual visual size would have been approximately 5.26 times higher than the values given above.

This selection of distances forms our Hypothesis $H2$. We wanted to investigate if the different range of distances used in two different experiments contributed to the contradictory results.

- **Tracking Method** $\in$ {Vision-based, Sensor-based} *within subjects*

  We have used the same tracking methods as described in Section 3 as a *within-subject* variable. Among four repetitions we have alternately used the tracking methods, i.e., two repetitions with vision-based tracking and two repetitions with sensor-based tracking. The order was counterbalanced among participants.

- **Repetition** $\in \{1, 2, 3, 4\}$ *within subjects*

  A set of 12 trials were randomly repeated four times for each participant, resulting in performing 48 different trials with one target object in each of them. We have used a $12 \times 12$ Latin square to randomize the trials.

**Dependent Variables**

Three quantitative variables were derived from the responses of the participants. Signed error (*SE*) was measured as the difference between the perceived distance and actual distance of virtual objects in meters. Hence, a positive *SE* indicates an overestimation of depth, while a negative *SE* indicates underestimation. We have also measured Absolute Error as $|SE|$. Time taken by the participants to respond on each of the trials was recorded.

### 4.4 Hypotheses

Before conducting the experiment we had the following hypotheses:

**H1** Our main motivation of designing photorealistic X-ray visualizations was to preserve the spatial relationship between foreground and background objects. Accordingly, we hypothesized that, the spatial arrangement *On-On* will have significantly better depth perception than the *On-Off* condition.

**H2** Results of our previous experiment showed $far-field$ distances are underestimated in an outdoor AR environment [7]; contradicting the results of another experiment where $medium-field$ distances were found to be overestimated in an outdoor AR environment [22]. So, we expected an overestimation of closer distances (up to 38.6 meters) and a gradual shift towards underestimation as distance was increased.

### 4.5 Results

We have prepared and analyzed the collected data using SPSS statistical package.

**Main Effects:**
Surprisingly, we did not find a main effect of Spatial Arrangement any of the dependent variables. There were main effects of Distance on Signed Error—F(11,264)=304.95, $p < .001$, $\eta_p^2$=0.93 and

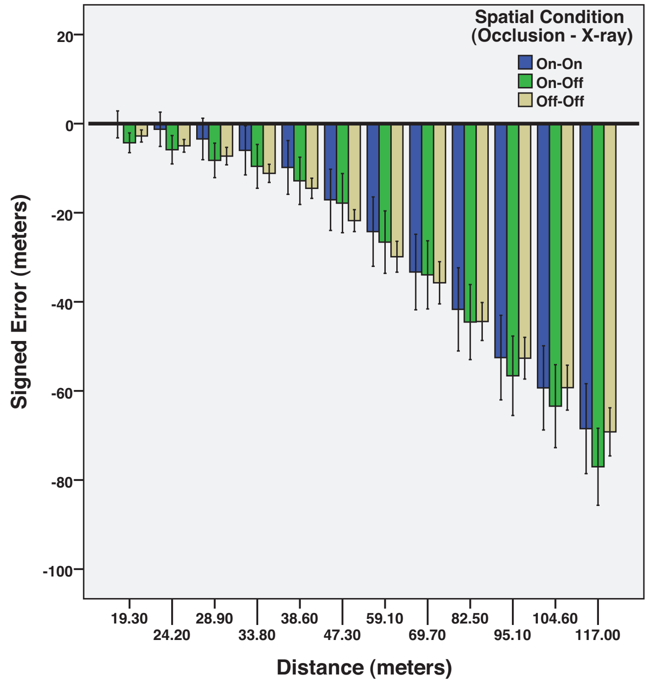



Figure 4: Signed Error shows a consistent underestimation of distances in all spatial conditions. Whiskers represent $\pm 95\%$ confidence interval and the thick black line represents veridical perception.

Absolute Error—F(11,264)=144.83, $p < .001$, $\eta_p^2$=0.86. Overall, with increasing distance Signed Error and Absolute Error increased (see Figure 4).

There was a main effect of Repetition on Response Time—F(1,24)=34.18, $p < .001$, $\eta_p^2$=0.59. Indicating a learning effect, the second repetition for each condition was significantly faster than the first repetition.

**Interaction Effects:**
An interaction effect of $Distance \times Repetition$ was found on Signed Error—F(11,264)=73.13, $p < .001$, $\eta_p^2$=0.14. Interestingly, signed error decreased in second repetition up to 47.3 meters, however, it started to increase from 59.1 meters onwards. There was a further interaction observed on Signed Error between $Distance \times Repetition \times Spatial\ Arrangement$—F(22,264)=4.72, $p < .001$, $\eta_p^2$=0.28. While On-On and On-Off conditions had consistently lower or similar signed error in second trials across all distances, Off-Off condition had considerably higher signed error in the second trials 47.3 meters onwards.

Similar interaction effects found on Absolute Error; first, $Distance \times Repetition$—F(11,264)=2.44, $p = 0.006$, $\eta_p^2$=0.09 and second, $Distance \times Repetition \times Spatial\ Arrangement$—F(22,264)=3.93, $p < .001$, $\eta_p^2$=0.25.

An interaction between $Distance \times Repetition$ was found on Response Time—F(11,264)=2.3, $p = 0.01$, $\eta_p^2$= 0.09. Though the second repetition was consistently faster than the first repetition, however, the difference was quite random.

### 4.6 Discussion

In this experiment, the comparison between Tracking Methods and Spatial Arrangements did not yield any significant difference on any of the dependent variables tested. Hence, we did not find any evidence in favor of accepting our hypotheses. We attribute this result to the difficulty of distance estimation using the small screen of the mobile phone. From the informal discussions with the participants, we found that the primary depth cue participants used was the relative change in the height of the pyramid in the visual field in successive trials. This is expected, as beyond 2 meters height in the visual field becomes the most important depth cue [6]. The change of height of the target pyramids were easily noticeable up to 36.6 meters, but after that point the change was not easily noticeable on the small screen and we noticed a sharp increase in the error (See Figure 4).

From the results of this experiment, it appears that the errors observed in tracking quality do not affect depth perception significantly. However, from the expressions of the participants we have observed some hints of discomfort when target pyramids jumped or started to move randomly. In some cases it caused a high response time. There is no interaction effect between the X-ray visualization and tracking error; also we did not find any significant effect to claim that not using X-ray visualization will effect depth perception for occluded objects; however, it is true that if we intend to go beyond the controlled experimental setup and unnatural target objects, a photorealistic rendering of the occluded region is essential which can not be achieved without X-ray visualizations.

However, in most of the cases, Off-Off condition performed better than other Spatial Arrangements. We have noticed that participants guessed the distance somewhat accurately up to 36.6 meters; this is expected from the explanations of Cutting: an average individual can accurately perceive distances up to around 30 meters (*action space*); but beyond that point (*vista space*) the distance perception gets increasingly compressed with distance [6].

Similar to our previous results in [7] and contradicting [22], we observed a consistent distance underestimation in an outdoor environment, even in medium field distances. Hence, our hypothesis H2 was refuted. Most importantly, we noticed that the depth compression in far-field distances is noticeably more than our previous experiment using a larger handheld display with a lower resolution [7]. To investigate the reason of this difference more deeply we have designed a second user study presented in Section 5, where we varied size and resolution of displays systematically in separate conditions.

## 5 Experiment 3: Effect of Display on Egocentric Depth Perception

In the last experiment we noticed a considerably higher depth compression compared to our previous experiment [7] using a larger display with lower resolution; consequently, we designed a within-subjects experiment to investigate the effect of size and the resolution of the display on depth perception. We were also interested to investigate if the X-ray visualization interacts with the display properties. As we did not notice any significant difference between the two tracking methods, in this study we have used the vision-based tracker.

### 5.1 Experimental Platform

We have used two different display devices. First, an iPhone 4s with 3.5” screen size and $960 \times 640$ (326 ppi) resolution. Second, an iPad3 with 9.7” screen and $2048 \times 1536$ (264 ppi) resolution (see Figure 1g). In this experiment we have used vision-based tracking method as participants subjectively preferred the same in Experiment 2.

### 5.2 Participants and Procedure

This was intentionally designed to be a within-subject experiment as we intended to evaluate all of the conditions using same participants eliminating any errors induced by separate participant groups. Total 12 participants (ages between 22 to 41 years) were recruited; among them six participants participated in the previous experiment at least two weeks ago.

Rest of the experimental procedure and used target object was similar to the first experiment. However, in this experiment there was no repetition in any condition. There were overall, six different conditions and we used a $6 \times 6$ Latin-square to randomize the presentation of the trials to the participants.

### 5.3 Variables

In this within-subjects experiment there were three independent and three dependent variables. Overall, there were $3\ (display\ configuration) \times 2\ (spatial\ arrangement) \times 12\ (distance) \times 12\ (participants)$= 864 data points.

**Independent Variables**

- **Display Configuration** (Combination of Size and Resolution) $\in$ {Small-Low, Big-Low, Big-High} *within subjects*

  We have used two different size and resolution of displays to create three different display configurations.

  **Small-Low:** This configuration is exactly the same configuration we used in the first experiment. The condition presented experimental trials on a 3.5” iPhone screen with $960 \times 640$ resolution. The horizontal visual size of the display was 9.52° (assuming that participants held the device at 45 cm away from their eyes).

  **Big-Low:** In this condition the experimental trials were presented on a 9.7” iPad3 screen with $960 \times 640$ resolution. The resolution was exactly the same with Small-Low condition. This display had a horizontal visual size of 24.7°.

  **Big-High:** In this condition the experimental trials were presented using the native iPad3 configuration, i.e. on a 9.7” screen with $2048 \times 1536$.

- **Spatial Arrangement** (Combination of Occlusion and X-ray Visualization) $\in$ {On-On, Off-Off} *within subjects*

  The same spatial arrangements as Experiment 2 were used. However, in this experiment we excluded On-Off condition, as we did not find any significant difference between the conditions.

- **Distance** $\in$ {19.3m (iPhone=1.86° vs. iPad=5.01°), 24.2m (1.48° vs. 4.08°), 28.9m (1.2° vs. 3.3°), 33.8m (1.04° vs. 2.86°), 38.6m (0.92° vs. 2.48°), 47.3m (0.74° vs. 2.1°), 59.1m (0.62° vs. 1.64°), 69.7m (0.5° vs. 1.46°), 82.5m (0.42° vs. 1.24°), 95.1m (0.36° vs. 1.06°),104.6m (0.34° vs. 0.94°), 117.0m (0.26° vs. 0.78°)} *within subjects*

  We have used exactly the same distances ranging from medium to far-field distances like the first experiment. However, depending on the size of the display, the virtual pyramids at the same distance had different visual sizes. Again, assuming a distance from eye to display of 45 cm, the pyramid at 19.3 m had a visual size of 1.86° on the iPhone and 5.01° on the iPad; values for all distances are given above within the parentheses after each distance. If the pyramids were real-world objects, then their visual size would have been approximately 5.26 times higher than on the iPhone and 1.87 times higher than on the iPad.

**Dependent Variables**

We have used exactly the same dependent variables like the first experiment—Signed Error (SE), Absolute Error (AE), and Response Time.

### 5.4 Hypotheses

Based on our previous experimental results we formulated the following hypotheses:

**H1** Both size and resolution of the display will have an effect on depth perception as bigger size and higher resolution provides improved viewing conditions.

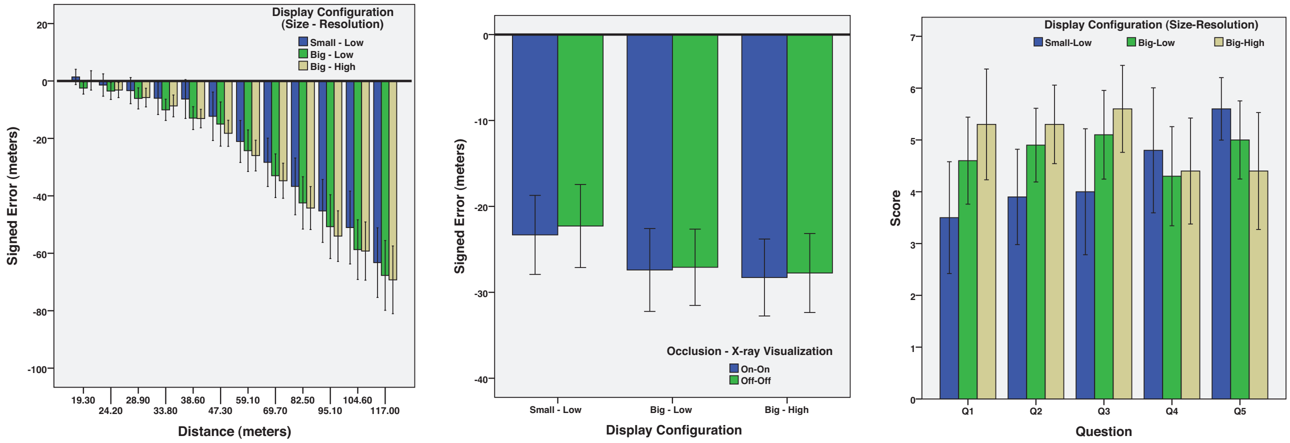
**H2** Following our previous experiments, we expected depth underestimation across all distances.

### 5.5 Results

The collected quantitative data was analyzed with a series of repeated measure ANOVAs using SPSS.

**Main Effects**

There was a main effect of display configuration on Signed




(a) Participants underestimated to a lesser amount using a small screen with low resolution.

(b) Underestimation was less in Off-Off spatial arrangement. This effect was significantly more in Small-Low display configuration.

(c) Subjectively Big-High condition was favored more than other conditions.

Figure 5: Results of Experiment 3: Effect of display size and resolution on egocentric depth perception. Whiskers represent $\pm 95\%$ confidence interval and thick horizontal black lines represent veridical perception.

Error—F(2,22)=4.29, $p = 0.027$, $\eta_p^2$=0.28. Interestingly, participants significantly underestimated distance least using Small-Low (iPhone) condition (see Figure 5(a)). Due to the large effect size we can assert that the differences in the degree of underestimation was indeed triggered by the display configuration. We have observed a significant main effect of distance on Signed Error – F(11,121)=88.76, $p < .001$, $\eta_p^2$=0.89 and Absolute Error – F(11,121)=78.06, $p < .001$, $\eta_p^2$=0.88. Expectedly, with increasing distance errors increased and accuracy decreased.

**Interaction Effects**
The only interaction effect we found was between *Display Configuration* $\times$ *Spatial Arrangement* on Signed Error—F(2,22)=9.82, $p = 0.03$, $\eta_p^2$=0.97 (see Figure 5(b)).

**Subjective Feedback**
After the experimental session we asked each participant to express their subjective feedback by marking each of the display conditions for following five attributes on a Likert scale of 1 to 7.(Q1) What was your confidence level of depth estimation? (Q2) What was your accuracy level of depth estimation? (Q3) How precisely have you identified the change of size of the faraway pyramids? (Q4) How much do you agree that you have underestimated the distance? (Q5) How much do you agree that you have faced difficulties to estimate distances of faraway pyramids?

Qualitative feedbacks was analyzed using a series of repeated measure ANOVAs (see Figure 5(c)). Participants were significantly more confident (Q1) while using Big-High display configuration followed by Big-Low and Small-Low—F(2,18)=12.85, $p < .001$, $\eta_p^2$=0.58. Participants assumed their accuracy level (Q2) to be significantly lowest in Small-Low condition—F(2,18)=11.32, $p = 0.001$, $\eta_p^2$=0.58. Similarly, they expected their precision in identifying the change of size of faraway objects (Q3) to be the lowest in Small-Low condition—F(2,18)=10.7, $p = 0.001$, $\eta_p^2$=0.54. Interestingly. all participants agreed strongly that they have underestimated the distance (Q4) in all conditions. Furthermore, participants agreed that they faced significantly more difficulties in estimating distances of faraway pyramids using Small-Low condition than Big-High condition—F(2,18)=6.85, $p = 0.006$, $\eta_p^2$=0.43.

### 5.6 Discussion

The primary finding of significance in this experiment is the similarity in depth perception using different display configurations, however, contradicting our expectations. Even the participants, through their subjective feedback, expressed a significant preference towards big screen and higher resolution. However, through objective measures we did not find any evidence of difference among different display configurations except for Signed Error. We observed that participants used both of the displays similarly, held them within 30 cm - 45 cm away from their eyes and did not move them intentionally.

We believe the difference on signed error was caused by the small size of the target pyramids on the small screen of an iPhone that prompted participants to imagine the pyramids to be further away than they were perceived using a bigger screen of an iPad. We have also noticed that the number of overestimations were most using a small screen. Two of the participants have consistently overestimated the distances using an iPhone (small-low configuration), whereas other participants were consistent with their estimation across all display configurations. It is also a point to note that, individual differences can attribute to differences in depth judgment experiments. We did not notice participants to move the displays a lot and the movement was minimal and similar in both of the displays. We attribute this to the missing motion parallax, caused by our rotation-only tracker.

We did not find any difference in resolution. Our chosen resolutions were based on market standards and was not controlled to vary largely. So the differences between resolutions was not enough to trigger different perception, hence, we did not find any significant difference. The fact that resolution has little effect on depth perception was also reported by [29].

Unlike the On-On condition, additional cues like Texture Gradient and Linear Perspective were present in Off-Off condition; however, we did not find any significant difference in depth perception. Hence, we assert that X-ray visualization having no effect on depth estimation is because height in the visual field was used mainly to interpret the distance. Participants have also confirmed this in informal discussions. Participants have informed that, instead of judging the distance based on the pyramid's height, they have measured the distance based on the relative change in height compared to the preceding trial; hence, height in the visual field was used like an ordinal cue.

Our second hypothesis was supported as we found a constant underestimation of distances across all conditions. This further supports the claim that in AR, like indoor locations, depth perception is underestimated in outdoor locations as well.

As participants expressed that they had significantly more difficulty in identifying the change of size and the egocentric depth of faraway pyramids, we expected a further study of exocentric and ordinal depth perception can lead to a comprehensive analysis of the perceptual effects. Hence, we have conducted a further experiment to investigate the exocentric and ordinal depth perception described in Section 6.

## 6 EXPERIMENT 4: EFFECT OF DISPLAY ON EXOCENTRIC AND ORDINAL DEPTH PERCEPTION

The previous experiment indicated that the size and resolution of a handheld display does not significantly affect egocentric depth perception. However, through a subjective questionnaire, participants reported that using small screen they were less confident in their depth judgment and they had higher difficulty in identifying the change of size of the far away pyramids.

In this experiment, we investigated the effect of display on exocentric and ordinal depth perception between two pyramids showed together. This type of depth judgment is required in real-world AR applications where multiple POIs are showed together.

### 6.1 Experimental Platform

The target objects and rest of experimental platform was identical to the Experiment 3 where egocentric depth perception was investigated (Section 5). However, in this experiment we excluded Big-Low display configuration. Similar to Experiment 3, we have used the vision-based tracker in this experiment.

### 6.2 Participants and Procedure

Same 12 participants from Experiment 3 were recruited for this *within-subject* experiment. However, there were at least one week's difference between two experiments for every participant. Participants were shown two pyramids of identical size together on the display. They had to identify which one of the two pyramids is closer by saying "Left", "Right", or "Equal". Then the had to verbally report the exocentric distance between the two pyramids (exocentric distance) to the experimenter. Participants were instructed to report the distance between the tips of the two pyramids.

Participants were allowed to take a break between the trials and the whole experimental session took about 25 minutes per participant.

### 6.3 Variables

In this within-subjects experiment there were three independent and four dependent variables. Overall, there were 2 $(display\ configuration) \times 2\ (spatial\ arrangement) \times 5\ (regions) \times 4\ (distance\ pair\ per\ region) 12\ (participants)$= 960 data points.

**Independent Variables**

- **Display Configuration** (Combination of Size and Resolution) $\in$ {Small-Low, Big-High} *within subjects*

  We had the same display configurations as the previous experiment. However, as we did not find any significant difference between Big-Low with two other configurations we removed it from this experiment. For the ease of discussion we will term small-low and big-high conditions as iPhone and iPad respectively in rest of the paper.

- **Spatial Arrangement** (Combination of Occlusion and X-ray Visualization) $\in$ {On-On, Off-Off} *within subjects*

  We had the identical spatial arrangements as the previous experiment—On-On ( Figure 1e) and Off-Off (Figure 1f). We counterbalanced the order of Spatial Arrangement.

- **Zone** $\in$ {1 to 5} *within subjects*

  In this experiment we presented the pair of target pyramids in five different zones. Each of those zones was 20 meters away from the starting distance of the previous zone. In each zone we have selected two distances, hence, in the whole experiment there were ten different distances where target pyramids were presented. The difference of egocentric distance between the two distances within each zone was 10 meters. The zones were – Zone 1 [30m (iPhone=1.15° vs. iPad=3.18°) - 40m (0.89° vs. 2.42°)], Zone 2 [50m (0.7° vs. 1.91°) - 60m (0.57° vs. 1.65°)], Zone 3 [70m (0.51° vs. 1.4°) - 80m (0.45° vs. 1.27°)], Zone 4 [90m (0.38° vs. 1.15°) -100m (0.35° vs. 1.02°)], and Zone 5 [110m (0.32° vs. 0.89°) - 120m (0.25° vs. 0.76°)]. If the pyramids were real-world objects, then their visual size would have been approximately 5.26 (iPhone) and 1.87 (iPad) times higher than they were on the screen.

  Within each zone we had four different stimuli. In two stimuli we presented the left and right pyramids at two different distance combinations. In the other two stimuli, we presented both of the pyramids at the same distances, however, one stimulus presented the pyramids at the far distance within the zone and the other at the near distance within the zone. Overall, in each session, 5 stimuli presented left pyramid closer, 5 stimuli presented right pyramid closer and 10 stimuli presented both pyramids at the equal distance to the participant. We have randomized the presentation of the stimuli.

**Dependent Variables**

We have used four dependent variables: 1. Signed Error, 2. Absolute Error, 3. Ordinal Error, and 4. Response Time. Signed Error and Absolute Error was calculated in the same way as previous experiments. The Ordinal Error is the error in the judgments of the relative depth—which one is closer?

### 6.4 Hypotheses

Based on our previous experimental results we formulated the following hypotheses:

**H1** iPad will have lower error in ordinal distance judgment than iPhone.

**H2** iPad will also have lower error in exocentric distance judgment.

### 6.5 Results

We have run a set of repeated measure ANOVAs to analyze the data.

**Main Effects:**

*Exocentric Perception-*
Like egocentric distances, exocentric distances were also underestimated on an average. There was a main effect of display configuration on Signed Error—F(1,11)=8.63, $p = 0.013$, $\eta_p^2$=0.44. Consistent to our earlier experiment, participants using an iPhone less underestimated the exocentric distance (see Figure 5a). Interestingly, a Chi-Square test indicated an significant main effect of Display on the number of overestimation and underestimation—$\chi^2$(1, $N = 960$)=55.9, $p < .001$. iPhone had significantly more number of overestimations than iPad (see Figure 6b).

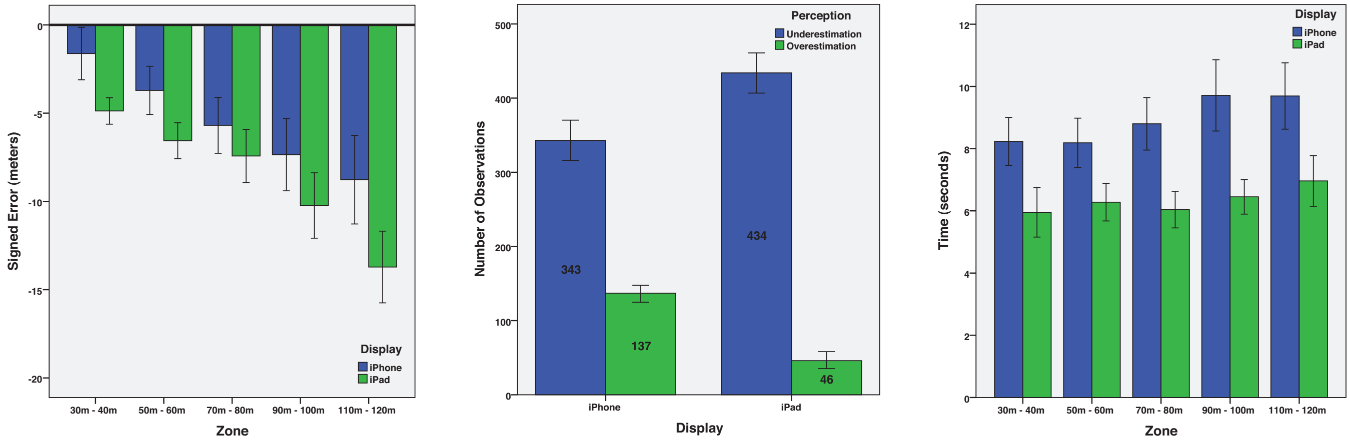
Though we did not find a significant effect of Display on Absolute Error (p=0.6); we have noticed a significant effect of Zone on Absolute Error—F(4,44)=11.6, $p < 0.0013$, $\eta_p^2$=0.51. Zone 5 had significantly higher error than zones 1 and 2.

In terms of Response Time, participants were significantly faster using an iPad than an iPhone—F(1,11)=10.77, $p = 0.007$, $\eta_p^2$=0.49 (see Figure 6c).

We did not find any significant effect of Spatial Arrangement on any of our dependent variables. There was no additional interaction effect between any of the variables tested.

*Ordinal Perception-*
A Chi-Square test indicated that, iPad (361 out of 460) had significantly more number of correct responses than iPhone (333 out of 460) in terms of ordinal depth perception—$\chi^2$(1, $N = 960$)=4.08, $p = .04$. Expectedly, Zone also had a significant effect on ordinal perception—$\chi^2$(4, $N = 960$)=32.94, $p < .001$. Errors consistently increased with distance.

**Subjective Feedback:**
After the experiment we asked participants to rate both of the displays on the following aspects on a Likert scale of 1 to 7. (Q1) What was your confidence level of ordinal depth judgment? (Q2) What was your confidence level of exocentric depth judgment? (Q3) How precisely have you identified the difference of size of the faraway




(a) Underestimation was less using an iPhone. The thick black line at 0 represents veridical perception.

(b) iPhone had significantly more number of overestimation of exocentric distances than iPad.

(c) Participants responded significantly faster using an iPad than an iPhone.

Figure 6: Results of Experiment 4: Effect of handheld displays on exocentric and ordinal depth perception. Whiskers represent ±95% confidence interval and thick horizontal black lines represent veridical perception.

pyramids? (Q4) How much do you agree that you have faced high task load?

We have conducted a set of paired two-tailed t-test on each of the questions. Participants expressed significantly higher confidence on both ordinal and exocentric depth perception when using an iPad—Q1. t(11)=-3.08, $p = .01$ and Q2. t(11)=-2.46, $p = .03$. Similarly, participants mentioned that they judged distance more precisely using an iPad than an iPhone—Q3. t(11)=-4.3, $p = .001$. iPad caused insignificantly higher task load.

### 6.6 Discussion

Our first hypothesis was supported as we found a significantly lower error in ordinal depth perception using an iPad than an iPhone. It was expected, as the difference between the height of the two pyramids was more identifiable on the big screen of an iPad than an iPhone.

Our second hypothesis was refuted as we did not find any significant difference in absolute error. However, like Experiment 3, we noticed less underestimation using an iPhone. We again assume that the smaller size of targets on an iPhone, subconsciously prompted participants to perceive the distance to be longer than on an iPad.

Similar to our earlier two experiments, again we did not find any effect of AR X-ray visualization on exocentric and ordinal depth perception. This results again indicates that, in AR environments, the height in the visual field is the primary depth cue in medium- to far-field distances. This finding is also encouraging, as this type of unnatural rendering of occluded objects does not hinder our depth perception; though, artificial environments do in general.

## 7 Conclusion and Future Work

We have presented, for the first time in the AR domain: (1) the influence of *handheld display* on depth perception and (2) the *exocentric* depth perception. In the detailed literature survey presented in Section 2, we have found only one previous depth perception experiment using a handheld display.

In Experiment 1, we presented a controlled benchmark of sensor-based and vision-based tracking methods. We did not notice any significant difference between the two.

Experiment 2 investigated the effect of distance-fields on human depth perception using a mobile phone. We have proven that distance is underestimated in medium to far-field distances in an outdoor AR environment; eliminating one of the possible factors of the contradiction noticed between one of our previous experiment [7] and an experiment by Livingston et al. [22]. Additionally we found, though tracking methods did not influence depth perception, participants subjectively expressed discomfort using sensor-based tracking when target objects jumped randomly. It indicates that accurate tracking is not very crucial for better depth perception in pedestrian applications like AR browsers.

In Experiment 3, we have investigated the effect of display size and resolution on egocentric depth perception. While we did not find a significant effect of resolution on depth perception; smaller display size caused less underestimation.

Experiment 4 investigated the effect of display properties on exocentric and ordinal depth perception. We have noticed that, bigger display caused significantly less ordinal errors. Ordinal error increased with increasing distance in both displays. In this study we have again noticed exocentric distance was underestimated throughout, and smaller display caused less underestimation.

In all of our user studies AR X-ray visualization did not affect depth perception. This result is encouraging, as our edge-overlay X-ray visualizations creates a certain amount of visual noise due to its edges; however, it did not affect perception of egocentric and exocentric distances in medium to far distances-fields.

In summary, our results have provided important insights on depth perception using handheld displays. First, bigger displays do not improve egocentric or exocentric depth perception; however, they significantly improve ordinal depth perception. Second, smaller displays cause less depth compression. Third, tracking methods and AR X-ray visualization do not influence depth perception in outdoor locations. Fourth, height in the visual field is the predominant depth cue for handheld AR applications.

More user studies are required to investigate the perceptual characteristics of handheld AR, particularly in outdoor locations. We noticed participants did not move the displays a lot. It will be interesting to investigate using a position tracker, hence enabling motion parallax. We assume that participants would move their displays more under these circumstances. We further assume that depth perception would be improved significantly, pointing out the necessity of position tracking for handheld AR. In the near future, lightweight and technically inferior HWDs will be introduced to the mass-market. While these HWDs will promote mass adoption of AR, however, they are expected to be of lower resolution and field of view than the commonly used HWDs in AR. More experimentation is required to understand how depth perception using this type of display may differ from currently used handheld displays. Similar studies to investigate differences in depth perception using optical see-through and video see-through HWDs are required in indoor locations. While we have used verbal reporting as the depth estimation protocol, there exist several other protocols. It will be interesting to investigate if using other protocols affect the depth judgment in outdoor AR environments. With more understanding of human perception in mobile AR and technical improvements, in the long-term, we aim to achieve ubiquitous AR [13].

## Acknowledgements

The authors wish to thank the voluntary participants of the evaluation, Michael Csongi for all his support, and reviewers for their valuable comments.

## References

[1] Apple. iphone4s. Last accessed on 28 February 2012. http://www.apple.com/au/iphone/specs.html.

[2] C. Arth, M. Klopschitz, G. Reitmayr, and D. Schmalstieg. Real-time self-localization from panoramic images on mobile devices. In *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '11, pages 37–46, Washington, DC, USA, 2011. IEEE Computer Society.

[3] B. Avery, C. Sandor, and B. Thomas. Improving spatial perception for augmented reality x-ray vision. In *Proceedings of the IEEE Virtual Reality Conference*, pages 79–82. IEEE, 2009.

[4] R. T. Azuma. A survey of augmented reality. *Presence*, 6:355–385, 1997.

[5] A. Clemens and D. Schmalstieg. Challenges of large-scale augmented reality on smartphones. *ISMAR 2011 Workshop: Enabling Large-Scale Outdoor Mixed Reality and Augmented Reality*, November 2011.

[6] J. E. Cutting. Reconceiving perceptual space. In H. Hecht, R. Schwartz, and M. Atherton, editors, *Looking into Pictures: An Interdisciplinary Approach to Pictorial Space*, chapter 11, pages 215–238. MIT Press, 2003.

[7] A. Dey, A. Cunningham, and C. Sandor. Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology*, pages 211–218, Hong Kong, China, November 2010.

[8] A. Dey, G. Jarvis, C. Sandor, A. Wibowo, and V.-V. Mattila. An evaluation of augmented reality x-ray vision for outdoor navigation. In *In Proceedings of International Conference on Artificial Reality and Telexistence*, pages 28–32, Osaka, Japan, November 2011.

[9] S. DiVerdi, J. Winter, and T. Höllerer. Envisor: Online environment map construction for mixed reality. In *Proc. VR 2008*, pages 19–26, Reno, Nevada, USA, March 8–12 2008.

[10] S. R. Ellis and B. M. Menges. Localization of virtual objects in the near visual field. *Human Factors*, 40(3):415–431, 1998.

[11] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban enviroment. In *Proc. ISWC'97*, pages 74–81, Cambridge, MA, USA, October 13–14 1997.

[12] J. M. Foley, N. P. Ribeiro-Filho, and J. A. D. Silva. Visual perception of extent and the geometry of visual space. *Vision Research*, 44(2):147 – 156, 2004.

[13] O. Hilliges, C. Sandor, and G. Klinker. Interactive prototyping for ubiquitous augmented reality user interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces*, IUI '06, pages 285–287, New York, NY, USA, 2006. ACM.

[14] I. Howard and B. Rogers. *Seeing in Depth*, volume 2. I Porteous, University of Torento Press, 2002.

[15] J. A. Jones, J. E. S. II, G. Singh, and S. R. Ellis. Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In R. McDonnell, S. J. Thorpe, S. N. Spencer, D. Gutierrez, and M. Giese, editors, *APGV*, pages 29–36. ACM, 2011.

[16] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Augmented Reality, 1999. (IWAR '99) Proceedings. 2nd IEEE and ACM International Workshop on*, pages 85 –94, 1999.

[17] H. Kim, G. Reitmayr, and W. Woo. IMAF: In-situ indoor modeling and annotation framework on mobile phones. *Personal Ubiquitous Computing*, 2012.

[18] G. Klein and D. W. Murray. Parallel tracking and mapping on a camera phone. In *Proceedings of the 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 83–86, 2009.

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

[20] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, 35(3):389 – 412, 1995.

[21] C. Levin and R. Haber. Visual angle as a determinant of perceived interobject distance. *Perception & Psychophysics*, 54:250–259, 1993.

[22] M. A. Livingston, Z. Ai, J. Swan, and H. S. Smallman. Indoor vs. outdoor depth perception for mobile augmented reality. In *VR '09: Proceedings of the IEEE Conference on Virtual Reality*, pages 55–62, Washington, DC, USA, 2009. IEEE Computer Society.

[23] M. A. Livingston, K. Karsch, Z. Ai, and LT Gregory O. Gibson. User interface design for military ar applications. *Virtual Reality*, 15(2–3):175–184, 2011.

[24] M. A. Livingston, J. Swan, J. L. Gabbard, T. H. Höllerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown. Resolving multiple occluded layers in augmented reality. In *ISMAR '03: Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 56, Washington, DC, USA, 2003. IEEE Computer Society.

[25] J. M. Loomis and J. M. Knapp. Visual perception of egocentric distance in real and virtual environments. In *L. J. Hettinger and M. W. Haas (Eds.), Virtual and Adaptive Environments*, number 11, pages 21–46. Mahwah NJ: Erlbaum, 2003.

[26] J. M. Loomis, J. W. Philbeck, and P. Zahorik. Dissociation between location and shape in visual space. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5):1202 – 1212, 2002.

[27] S. Lovegrove and A. J. Davison. Real-time spherical mosaicing using whole image alignment. In *Proc. ECCV 2010*, volume 6313, pages 73–86, Heraklion, Crete, Greece, Sept 5–11 2010.

[28] J. McCandless, S. R. Ellis, and B. D. Adelstein. Localization of a time-delayed, monocular virtual object superimposed on a real environment. *Presence*, 9(1):15–24, 2000.

[29] K. mikael J-aro and L. Kjelldahl. Effects of image resolution on depth perception in stereo and non-stereo images, 1997.

[30] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *In Proceedings of the SPIE Conference on Telemanipulator and Telepresence Technologies*, pages 282–292, 1994.

[31] Mobilizy. Wikitude. Last accessed on 28 February 2012. http://www.wikitude.com/.

[32] S. Nagata. How to reinforce perception of depth in single two-dimensional pictures. In *Pictorial Communication in Virtual and Real Environments*, chapter 35, pages 527–545. Taylor & Francis, Inc., 1991.

[33] B. Pollock, M. Burton, J. Kelly, S. Gilbert, and E. Winer. The right view from the wrong location: Depth perception in stereoscopic multi-user virtual environments. *Visualization and Computer Graphics, IEEE Transactions on*, 18(4):581 –588, april 2012.

[34] J. Predebon. Spatial judgments of exocentric extents in an open-field situation: Familiar versus unfamiliar size. *Attention, Perception, & Psychophysics*, 50:361–366, 1991. 10.3758/BF03212228.

[35] G. Reitmayr and T. W. Drummond. Going out: Robust modelbased tracking for outdoor augmented reality. In *Proceedings of 5th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR*, pages 109–118, 2006.

[36] C. Sandor, A. Cunningham, A. Dey, and V.-V. Mattila. An augmented reality x-ray system based on visual saliency. In *In Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pages 27–36, Seoul, Korea, October 2010.

[37] G. Singh, J. E. S. II, J. A. Jones, and S. R. Ellis. Depth judgments by reaching and matching in near-field augmented reality. In S. Coquillart, S. Feiner, and K. Kiyokawa, editors, *VR*, pages 165–166. IEEE, 2012.

[38] SPRXmobile. Layar reality browser. Last accessed on 28 February 2012. http://www.layar.com/.

[39] J. I. Swan, M. A. Livingston, H. S. Smallman, D. Brown, Y. Baillot, J. L. Gabbard, and D. Hix. A perceptual matching technique for depth judgments in optical, see-through augmented reality. In *VR '06: Proceedings of the IEEE Conference on Virtual Reality*, pages 19–26, Washington, DC, USA, 2006. IEEE Computer Society.

[40] T. Tsuda, H. Yamamoto, Y. Kameda, and Y. Ohta. Visualization methods for outdoor see-through vision. In *ICAT '05: Proceedings of the International Conference on Augmented Tele-Existence*, pages 62–69, New York, NY, USA, 2005. ACM.

[41] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg. Real-time panoramic mapping and tracking on mobile phones. In *Proceedings of the IEEE Virtual Reality Conference*. IEEE, 2010.

[42] D. Wagner, G. Reitmayr, A. Mulloni, T. W. Drummond, and D. Schmalstieg. Real time detection and tracking for augmented reality on mobile phones. *IEEE Transactions on Visualization and Computer Graphics*, 16(3):355–368, May/June 2010.

[43] C. Ware. *Visual Thinking: for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2008.

[44] J. Wither and T. Hollerer. Pictorial depth cues for outdoor augmented reality. In *ISWC '05: Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, pages 92–99, Washington, DC, USA, 2005. IEEE Computer Society.

[45] F. Zhou, H. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ismar. In *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*, pages 193 –202, sept. 2008.