Is Autostereoscopy Useful for Handheld AR?

Frederic Kerber[1], Pascal Lessel[1], Michael Mauderer[2], Florian Daiber[1], Antti Oulasvirta[3], Antonio Krüger[1]

[1]German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany - firstname.lastname@dfki.de
[2]School of Computer Science, University of St Andrews, St Andrews, United Kingdom - mm285@st-andrews.ac.uk
[3]Max Planck Institute for Informatics and Saarland University, Saarbrücken, Germany - oantti@mpi-inf.mpg.de

ABSTRACT

Some recent mobile devices have autostereoscopic displays that enable users to perceive stereoscopic 3D without lenses or filters. This might be used to improve depth discrimination of objects overlaid to a camera viewfinder in augmented reality (AR). However, it is not known if autostereoscopy is useful in the viewing conditions typical to mobile AR. This paper investigates the use of autostereoscopic displays in an psychophysical experiment with twelve participants using a state-of-the-art commercial device. The main finding is that stereoscopy has a negligible if any effect on a small screen, even in favorable viewing conditions. Instead, the traditional depth cues, in particular object size, drive depth discrimination.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – Artificial, augmented, and virtual realities

Keywords

Autostereoscopy; mobile devices; depth discrimination; empirical and quantitative user study; augmented reality.

1. INTRODUCTION

The two well-known display technologies possible for mobile augmented reality (AR) are head-mounted displays (HMD) and video see-through displays. In contrast to HMDs that provide mono- as well as stereoscopic projection, systems that rely on video see-through displays have been based on monoscopic displays. Anecdotal evidence suggests that users have difficulties in assessing the distance of virtual objects on magic lens displays.

A currently emerging class of handheld devices is equipped with autostereoscopic displays, which in principle would enhance users' 3D perception without additional user instrumentation (i.e. shutter or anaglyph glasses). These displays allow for the presentation of objects in different parallaxes; that is, negative (NEG-P, in front of the screen), zero (at the screen level), and positive (POS-P, behind the screen) parallax, resulting in different stereoscopic effects. Autostereoscopic displays are known to enhance depth discrimination in large displays like televisions [6]. However, it is not known if the positive effect is reproducable for handheld viewing conditions. Moreover, the mobile case involves a drastically smaller display size. Currently available mobile autostereoscopic devices use the parallax barrier technique[1] which has constraints on small displays, such as limited viewing range and field of view.

This paper contributes an experiment that investigates depth discrimination on a state-of-the-art commercial autostereoscopic mobile device. Virtual objects are overlaid on a real-world scene on a camera viewfinder and the participants have to distinguish which one is closest to them. We address the following two questions: (I) Does autostereoscopy improve users' depth discrimination ability, or do they rely more on monoscopic cues such as object size? (II) Does the presentation of virtual objects in different parallaxes influence the depth discrimination ability?

The binocular parallax depth cue might help in the discrimination of objects especially in a densely cluttered environment. These questions are of particular interest for AR settings where virtual objects with different sizes are used. The main goal of the study is to measure the effects of stereoscopy in a mobile context and in this way exclude the influence of other depth cues. Thus, we investigate only binocular parallax and object size in the experiment. Our experiment considers both negative and positive parallax. Stereoscopy is expected to improve spatial perception and guidance of virtual objects in AR scenarios. However, in stereoscopic mobile AR the see-through metaphor only holds for the POS-P case because the real environment in the camera view lies behind the device. Nonetheless, the NEG-P condition is also taken into consideration since it can even be used in an AR scenario (e.g. for controls).

2. RELATED WORK

Depth interpretation is a common problem in AR applications and creating a perceptually correct augmentation is still a challenge [4]. Few research has investigated the use of autostereoscopic mobile devices for AR. Nonetheless, some work focuses on aspects that are of relevance here.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MUM '13, Dec 02-05 2013, Lulea, Sweden

---

[1] see e.g. `http://bit.ly/14BLi87`, last accessed Oct 28nd, '13

Dey et al. [2] investigated depth perception on handheld devices with different screen sizes (i.e. iPad and iPhone). None of these devices have autostereoscopic capabilities. The results of several experiments show that there is no significant effect of screen resolution on depth perception, but there is an effect on distance estimation.

Huhtala et al. [3] investigated whether autostereoscopy could help users in a selection task where relevant parts were highlighted. Four conditions were tested, two of them involving autostereoscopic cues. The results did not show that using stereoscopy alone improved the performance, but the combination with a second visual cue performed better.

The work of Mikkola et al. [5] considers the importance of different depth cues on a mobile autostereoscopic display. Participants were presented with several virtual balls that had been placed at different depths on a virtual background. For different depth cues, the participants had to decide which of the balls was at the same depth as a reference object. The results show that the stereoscopic depth cues outperform the monocular ones in accuracy and speed of depth estimation.

Recently, Broy et al. investigated depth discrimination on stereoscopic displays [1]. They motivate their work for the field of autostereoscopic display in the automotive domain. However, for their evaluation they use a steroscopic display and shutter glasses. Another difference from our work was that they only considered a virtual scene and no AR setting for evaluating the depth discrimination ability. Furthermore, no comparison of the steroscopic vs. non-stereoscopic condition was done. Their findings are somewhat contradictory to our results, which we will address in our discussion.

3. EXPERIMENT

Building on methodology from psychophysics [7], we designed an experiment with two discrimination tasks. Realizing the limitations of present-day autostereoscopic displays, we calibrated the viewing conditions to be as close to ideal as possible. A chin rest was used to ensure that the participants had a consistently stereoscopic effect throughout the whole experiment. We noticed in a pilot study that, without such constraints, users would — lacking previous experience with such devices — intuitivly hold the device at angles and distances unfavorable to the autostereoscopic display. Our data thus is based on the best-case scenario.

3.1 Participants

Twelve participants (two female, age between 21 and 34, mean $M = 25.3$) were recruited. All of them were informed about the aim of the study and the procedure. The participants were students (75%) and researchers (25%). They were invited for 2-3 sessions to ensure that eyestrain would not affect the results. A session lasted 47.9 minutes on average. The first session allowed for completion of Task I and the first half of Task II. The second and third sessions allowed for the completion of the remaining part of Task II.

Every participant had normal or corrected-to-normal vision. Ten of them reported prior experience with stereoscopic effects (e.g. 3D cinema) and four reported prior experience with autostereoscopic devices. To ensure that all participants are capable of perceiving 3D, we tested their stereo vision capabilites in a pre-test. In this test several cubes of fixed size were subsequently shown on a white background, either in NEG-P or POS-P. Participants had to state whether they had the impression of cubes floating in front (NEG-P) or behind the display (POS-P). The results showed that none of the participants had any severe problems with stereo vision (success rate between 85% and 100%, $M = 94.2\%$).

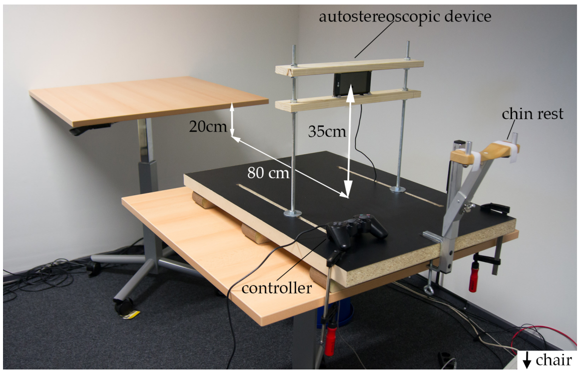

Figure 1: Experimental setup: chin rest, fixed autostereoscopic device and background scenery.

3.2 Apparatus

We considered the two main autostereoscopic smartphones that are currently available on the market: a HTC Evo 3D and a LG Optimus 3D Max. Because of incompatibilities in HTC's 3D SDK, the device from LG was chosen. The smartphone's dimensions are $126.8 \times 67.4 \times 9.6\ mm$ with a 4.3-inch screen, having a resolution of 480×800 pixels.

The device was fixed in a frame as depicted in Figure 1. With this setup, we ensured a constant distance from the device to the scene of interest, as well as a constant distance between viewer and device, for a consistent 3D effect. To compensate for individual differences in head size (i.e. length between chin and eyes) and body height, the chin rest, the chair and both tables could be adjusted for height. At the beginning of each task, it was ensured that the participants' eyes were at the right height. Participants were asked to maintain a constant seated position during the task. For all situations, we ensured that the relative difference in height between the two tables remained constant. The device was always mounted at the same height on the first table. Furthermore, it was ensured that the same illumination conditions were used for all participants.

The device's camera image showing a real-world table was augmented by two virtual cubes floating 200 mm above the table (see Figure 2). Artifical, clean background AR scenery close to a virtual reality (VR) setting was chosen as the best-case scenario. All cubes were presented with the same texture. No lighting effects were used, to avoid introducing additional visual cues. Again, in the experimental setup, the real and the virtual space had to be carefully integrated (i.e. the disparities between virtual space and camera space) to ensure that no other influences affected the study. This also constrained the available space where objects could be placed at a reasonable size and without touching the display's borders.

Figure 2: Autostereoscopically projected objects augmented on camera view.

3.3 Task I: Object sizes

We investigated the influence of the objects' sizes on depth discrimination. The camera image and the cubes were always shown autostereoscopically. Two cubes were used to ease depth discrimination. Cube size and parallax were considered as independent variables within subjects. We uniformly varied whether cubes were shown big/small and in NEG-P/POS-P, resulting in 4 different conditions which were presented 5 times each in random order. The cubes' sizes were adjusted such that the small cubes in NEG-P have the same apparent size as the big cubes in POS-P and vice versa. Cubes in NEG-P were placed at a depth of 700 mm, those in POS-P at 1700 mm (referring to the cubes' front face, measured from the camera's position towards the background wall). Participants were told that both cubes are placed at the same depth. They were asked to decide, as the dependent variable, whether the cubes are shown in NEG-P or POS-P and report their choice verbally.

3.4 Task II: Autostereoscopy and Parallax

The second task investigated the effect of autostereoscopy as well as the parallax on the depth discrimination capability. The depth of the two virtual cubes were varied and the participants had to decide which of the two cubes was closer to them (see Figure 2). With the help of an adaptive stair-casing procedure, we determined the required minimal depth distance between the two virtual cubes to be able to discriminate them (dependent variable). We considered three independent variables within subjects:

Autostereoscopy (*On*, *Off*): To measure the effect of the autostereoscopic cue, we integrated two conditions. We displayed the camera image as well as the augmented cubes with stereoscopic effects (*On* condition) and without (*Off* condition). In the latter, no stereoscopic effects were enabled: Only one of the stereo camera images was used (randomly chosen) together with the virtual cubes.

Size (*Randomized*, *Fixed*): To test the influence of the size cue, we varied all the cubes' edge lengths in the *Randomized* condition uniformly distributed between 75 mm and 100 mm for every stimulus presentation. In the *Fixed* condition both cubes had a constant size of $87.5 \times 87.5 \times 87.5\ mm^3$. Additionally, the sizes were adjusted perspectively correct regarding displayed depth in both conditions.

Object depth (*700*, *800*, *900*, *1400*, *1500*, *1600*, *1700 mm*): To check the influence of the cubes' depths (referring to their front face, measured from the camera's position towards the background wall) we considered seven different conditions, the first three in NEG-P, the latter four in POS-P.

We counterbalanced the first two independent variables (autostereoscopy and size) via a balanced Latin square and randomized the order of the possible cube depths.

Parallax	Size	Mean error rate	t-Tests
NEG-P	Big	6.67%	$t(59) = 13.34$, $p < .001$
NEG-P	Small	45.00%	$t(59) = 0.774$, $p > 1$
POS-P	Big	83.33%	$t(59) = -6.87$, $p < .001$
POS-P	Small	13.33%	$t(59) = 8.29$, $p < .001$

Table 1: Conditions and given answers of Task I. Single-sample t-tests against .5 to determine deviation from chance. p-values are Bonferroni corrected.

As study design an adaptive stair-casing procedure was used: The presented stimulus remains the same until a discrimination capability can be assumed or rejected with a certain confidence. If the participant can discriminate the stimulus, the next is presented with reduced intensity. Otherwise, it is presented with a higher intensity. The goal of this procedure is to find the minimal intensity the participant is able to discriminate. Instead of using a fixed number of targets, we decided to use a PEST procedure [7] for this, as it has the advantage of adjusting the change in intensity based on the prior performance to achieve a faster convergence towards the final intensity level. In our case, the stimulus intensity maps to the depth distance of the cubes' front faces to each other and additionally, after one completed PEST procedure (i.e. change in stimulus intensity $\leq 1\ mm$), a new PEST procedure was started with a different object depth. For every stimulus presentation, one randomly chosen cube was displaced in depth accordingly. Between every two steps, a fixation crosshair (a black cross on a background similar to the wall in the camera image) was shown for 500 ms in order to force accomomdation switches. Participants were instructed to decide which of the two shown cubes was closer to them and then report their choice by pressing one of the corresponding shoulder buttons on a Playstation 2 controller even if they were unsure about their decision (two-alternative forced-choice).

4. RESULTS

In the following, the results of Task I and II are reported.

4.1 Task I: Object Size

In Task I the number of correct answers varied between 50% and 75% ($M = 62.9\%$). Table 1 shows the distribution of answers in the different conditions. Participants mainly judged cubes to be in NEG-P if displayed big and to be in POS-P if displayed small. An interesting finding is that participants were able to discriminate in all conditions except for NEG-P small. Even though the error rate for POS-P big is very high, it is significantly *above* chance (see Table 1), indicating a consistently wrong classification. The size variable and the depth that was assumed by the participants are strongly correlated (Pearson's $r(478) = 0.56$, $p < 0.01$), whereas the real depth and the assumed depth are only slightly correlated (Pearson's $r(478) = 0.26$, $p < 0.01$).

4.2 Task II: Autostereoscopy and Parallax

Figure 3 shows the results of Task II. The x-axis shows the seven object depths; the y-axis illustrates the mean minimal depth distance between the cubes' front faces to enable a discrimination. The four conditions (autostereoscopy *On*/*Off* and size *Fixed*/*Randomized*) are shown separately.

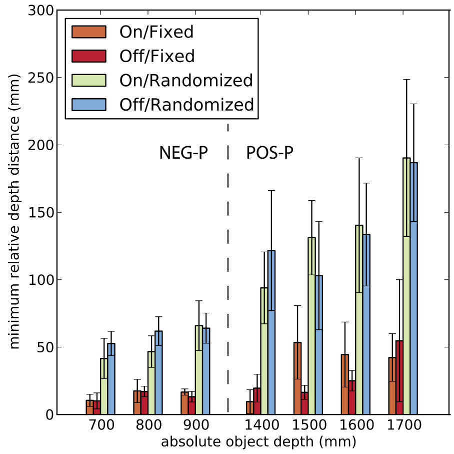

Figure 3: Results of Task II showing mean relative depth distances required for discrimination of virtual objects at different depths for the independent variables autostereoscopy and size (error bars indicate 95% confidence intervalls).

To investigate the effect of the different conditions on the minimal depth distance needed between the cubes' front faces to enable a discrimination, a $2 \times 2 \times 7$ repeated-measure ANOVA was performed. Where Mauchly's test indicated that the assumption of sphericity had been violated, Greenhouse-Geisser correction was applied. Results indicate significant main effects for object depth ($F(3.08, 33.88) = 30.07$, $p < .001$, $\eta_p^2 = .73$) and for size ($F(1, 11) = 289.42$, $p < .001$, $\eta_p^2 = .96$) with static size having a lower minimal distance ($N = 168$, $M = 25.06$, $SD = 4.06$) than random size ($N = 168$, $M = 102.40$, $SD = 4.66$). Post-hoc contrasts for object depth reveal a significant quadratic trend ($F(1, 11) = 5.63$, $p = .037$, $\eta_p^2 = .59$)). No significant main effect for stereo was found ($F(1, 11) = .10$, $p = .762$, $\eta_p^2 = .01$). In addition, there is a significant interaction for size and depth ($F(3.45, 38.00) = 15.79$, $p < .001$, $\eta_p^2 = .59$).

We draw the following conclusions about the data from Task II: First, the minimal required depth increases with increasing absolute depth of the virtual objects to be compared. Second, no significant differences in the comparison of the autostereoscopic conditions (*On*/*Off*) could be found. In other words, autostereoscopy did not change performance in the task. Third, it was also easier for the participants to discriminate objects in the *Fixed* size conditions.

5. DISCUSSION AND OUTLOOK

We investigated human depth discrimination in AR on a small mobile device using a commercial autostereoscopic device. We learned that in these conditions, stereoscopy has a negligible effect on the users' ability to distinguish the depth of virtual objects imposed on a real scene. In Task II, we would have expected a positive effect of autostereoscopy in the *Fixed* size condition, but this did not appear even in the viewing conditions that were calibrated to the user and the device. We attribute the lack of an effect to the competing cues in the overlaid objects (figure) and the background (camera viewfinder image). We piloted previously with a virtual-only scene and observed that there can be a positive effect of autostereoscopy. Our results suggest that the positive effect disappears in the AR condition when the VR objects are seen superimposed on the viewfinder's image. For the present study, we worked hard to adjust the effect to match the provided camera image to ensure that no perceptual mismatch occured. We believe that it is hard for users to fuse the two representations based on the autostereoscopic cues, and they instead rely on the monoscopic cues. This hypothesis warrants further study. The study also sheds light on some underlying perceptual factors. The results of the first task show that people rely more on the object size cue than on the autostereoscopic cue. In Task II, we found that the capability of discriminating depths in NEG-P is better than in POS-P. This finding is consistent with the first task, in which fewer errors were produced in NEG-P. Nonetheless, it contradicts the results of Broy et al. [1]. As the chosen AR background is close to a VR setting, we anticipate that this is due to the differences in display technology (autostereoscopic vs. shutter glasses) and size (4.3 inches vs. 17 inches). But this is a topic for further research.

To conclude, present-day autostereoscopic displays are not superior to monoscopic displays with regards to mobile AR. In other words, the autostereoscopic cue should not be relied on as a primary cue for depth discrimination. Instead, the size cue appears to be a dominant over the autostereoscopic cue. Hence, placing large objects in the back should be done carefully as it can lead to false depth perception. The results also indicate that virtual objects of known size can be arranged closer to each other. Users will also benefit from displaying objects in NEG-P rather than POS-P.

In future work, it is important to study factors affecting depth discrimination in conditions that involve free user movement in the scene. We hypothesize that, should there be any effect, it will be spatially constrained. Spatial constraining is counter-productive in mobile use where the viewing angle and distance change dynamically. Further experiments should investigate other depth cues as well as scenarios that integrate and augment real and virtual objects.

6. REFERENCES

[1] N. Broy, F. Alt, S. Schneegass, N. Henze, and A. Schmidt. Perceiving layered information on 3D displays using binocular disparity. In *Proc. PerDis*, pages 61–66. ACM, 2013.

[2] A. Dey, G. Jarvis, C. Sandor, and G. Reitmayr. Tablet versus phone: Depth perception in handheld augmented reality. In *Proc. ISMAR*, pages 187–196. IEEE, 2012.

[3] J. Huhtala, M. Karukka, M. Salmimaa, and J. Häkkilä. Evaluating depth illusion as method of adding emphasis in autostereoscopic mobile displays. In *Proc. MobileHCI*, pages 357–360. ACM, 2011.

[4] E. Kruijff, J. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *Proc. ISMAR*, pages 3–12. IEEE, 2010.

[5] M. Mikkola, A. Boev, and A. Gotchev. Relative importance of depth cues on portable autostereoscopic display. In *Proc. workshop MoViD*, pages 63–68. ACM, 2010.

[6] P. Surman. Stereoscopic and autostereoscopic displays. In C. Zhu, Y. Zhao, L. Yu, and M. Tanimoto, editors, *3D-TV System with Depth-Image-Based Rendering*, pages 375–411. Springer, 2013.

[7] M. M. Taylor and C. D. Creelman. PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 41:782–787, 1967.