Improving Depth Perception for Hand-held Augmented Reality using Autostereoscopic Displays

Matthias Berning†  ‡  Daniel Kleinert †  Till Riedel‡  Michael Beigl§
Karlsruhe Institute of Technology (KIT), TECO
Karlsruhe, Germany

ABSTRACT
Displaying three-dimensional content on a flat display is bound to reduce the impression of depth, particularly for mobile video see-through augmented reality. Several applications in this domain can benefit from accurate depth perception, especially if there are contradictory depth cues, like occlusion in an x-ray visualization. The use of stereoscopy for this effect is already prevalent in head-mounted displays, but the there is little research on the applicability for hand-held augmented reality. We have implemented such a prototype using an off-the-shelf smartphone equipped with a stereo camera and an autostereoscopic display. The system achieves interactive frame rates by restricting the tracking to the left image and camera and an autostereoscopic display. The system achieves interactive frame rates by restricting the tracking to the left image and camera and an autostereoscopic display. The systems achieves interactive frame rates by restricting the tracking to the left image and camera and an autostereoscopic display. The results show that our system can significantly improve depth positioning accuracy, especially in scenes that lack other cues.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Ergonomics

1 INTRODUCTION
Depth distortion is still one of the most common perceptual problems found in Augmented Reality (AR) today [7]. It describes the issue that users cannot correctly identify the spatial relation between objects based on their viewpoint, which is especially the case for the combination of real and virtual objects. Depending on the application this information can be crucial for a good performance, e.g. in maintenance task, manuals or when visualizing occluded objects [15].

There is already a large body of research in the area of depth perception for AR, which is focused on head-mounted displays [17, 9, 16, 4]. This is viable for professional applications, but due to commercial reasons, the current platform of choice for consumer oriented AR, are mobile hand-held devices. Since they are limited to video see-through AR and offer only a small field-of-view, these devices are even more prone to depth distortion.

The availability of smartphones featuring an autostereoscopic display could enable the use of binocular depth cues in this area. The displays promise a high degree of immersion and increased spatial awareness through a single screen without the need for additional glasses worn be the user (i.e. polarized or shutter). In combination with the two cameras on the back of the device, it is possible to realize true stereoscopic AR.

Our contribution in this work is an experimental evaluation on the effect of stereopsis for depth perception in mobile hand-held AR. Different virtual objects were embedded in a real scene and displayed to the user. The participants in our study had to align them accurately with a real object. The goals were a) to verify if the autostereoscopic display can help an user to understand the spatial relation between real and virtual objects and b) to compare the effect to other depth cues.

2 DEPTH PERCEPTION AND STEREOSCOPIC DISPLAYS
According to Cutting, [2] the relative importance of different depth cues is determined by the distance of the objects to the user. There are three different areas: In personal space, from 0-2m directly in front of the observer, binocular disparity provides the most accurate depth judgments. It is the most important depth cue provided by stereoscopic vision and particularly useful to resolve ambiguities created by other perceptual cues. Kytö et al. [8] list some of its special benefits for AR, e.g. the layering of augmentations to increase the information density.

Current autostereoscopic displays are either of the lenticular sheet or the parallax barrier type [12]. Both have in common, that a normal display is used and the images for the left and right eye are arranged in interleaved columns. By applying a sheet to structure the emitted light, each image is only visible from a specific angle which should be the corresponding eye of the user. Due to this arrangement, stereopsis is only achieved in a specific distance and orientation to the display, which is one of the major drawbacks. Another problem is the conflict between accommodation and convergence, which are directly linked in normal vision. When consuming binocular content on a flat screen, the relative orientation of the eyes will change (convergence) but the focus has to stay fixed on the screen (accommodation). Nevertheless, recent studies have shown that these devices can improve the user experience when consuming content on a hand-held device [13]. To this date, off-the-shelf smartphones are only available with parallax barrier type displays. They structure light by superimposing fine vertical lines onto the display, blocking every second column for a specific point of view. Through the use of liquid-crystal arrays for this barrier, the stereoscopic function can be controlled dynamically.

3 RELATED WORK
Improving the perception of depth is a widely researched topic in VR and AR, although most of it is focused on HMDs. The work of Swan et al. [17] is exemplary for the conducted studies and is listed here explicitly because of the survey on the field included in the publication. They conducted two experiments to investigate the effects of stereoscopic vision on depth judgment in head-worn AR. Both showed a main effect of stereoscopy on depth judgment resulting in greater accuracy.

Dey et al. [3] looked at the effect of different hand-held screen sizes and resolutions on depth-perception. In several AR test scenarios, they found a significant effect of resolution on distance estimation. Autostereoscopic displays were not part of their evaluation.
The authors of [1] conducted a Wizard-of-Oz study, using a cutout phone to simulate the effects of binocular vision on positioning accuracy of a hand behind the device. They show that performance was worse using a monocular setting. A functional prototype using video see-through could not be tested.

The Nintendo 3DS\(^1\) is a hand-held gaming console that supports stereoscopic AR. There are some games that use this feature, but the development on this platform is restricted by the manufacturer. Structured evaluations on the effects of depth-perception were not possible with the available software.

The work of Kerber et al. [6] is the most relevant for our own experiments. Similar to our work, they conducted an experiment to determine the depth discrimination ability of participants using a smartphone with an autostereoscopic display. For this purpose, two virtual textured cubes with varying size were displayed above a real table and the subjects had to decide which one was closer. In their analysis, they found no significant effect of the stereoscopic condition on depth perception. Unfortunately, the paper does not clarify how they handled the automatic disparity remapping performed by the phone. As detailed in the next chapter, this can have a severe effect on the perceived AR experience. Another aspect that is not addressed is the alignment of real and virtual objects, since the tested scenario was deliberately chosen to be very close to VR.

In our opinion, based on previous research, it is especially this interaction that could benefit the most from stereoscopic vision, which is why this is the focus of our work.

Mikkola et al. [13] investigated the effect of small autostereoscopic displays on depth perception in static scenes. They found a significant improvement of accuracy and speed of depth judgements when introducing binocular disparity. The addition of secondary depth cues as shadowing, texture gradient or depth blur had no measurable effect on this result.

4 Prototype Implementation

We implemented our prototype on the off-the-shelf smartphone LG Optimus 3D Max (P720) featuring a stereo camera and an autostereoscopic display of the parallax barrier type. The system is based on Android and runs the latest official firmware for this device with version 2.3 of the mobile operating system. Processing resources are limited by the dual-core ARM processor, rated at 1.2GHz and 1GB of RAM. The 4.3in display has a native resolution of 800 × 480 in landscape orientation, which is halved horizontally when enabling the parallax barrier. The camera pair has a stereo basis of 24mm and a resolution of 5MP each, although we are only using a much lower resolution for tracking. Both hardware units are controlled via the proprietary Real3D API provided by LG, which extends the Android framework to support 3D user interfaces.

The software is based on AndAR\(^2\), which provides a solid architecture including the necessary modules for camera control, tracking and rendering, mapped to the Android OS. The basic frame processing is depicted in Figure 1. We extended the framework in several aspects to increase performance, while maintaining or even reducing resource consumption. This includes upgrading the tracking library to ARToolKitPlus\(^3\) [18] and replacing the renderer with Rajawali\(^4\). To maximize throughput for the stereoscopic mode, we limit tracking to the left video stream and calculate the transformation for the virtual objects in the right part based on the camera offset. In addition to the lower resource consumption, this also prevents tracking errors from interfering with the stereoscopic effect.

We had to find the optimal configuration for the LG Real3D API, which provides fine grained control over the camera and display hardware. Camera frames are set to be placed in top-bottom arrangement in memory, meaning the left frame is located before the right, allowing for sequential access to each one. By using the YUV image format, the luminance component can be separated easily and is forwarded to the tracker. The used tracking resolution is 640 × 480. Camera calibration from the OpenCV package produced the intrinsic camera parameters, as well as the extrinsic transformation matrix \(M_E\) from the left to the right camera image.

After our first trials provided an unpleasant viewing experience due to unsatisfactory alignment, we identified the automatic disparity remapping of the phone to be responsible. Disparity remapping is used in stereoscopic video recordings to warp the source images to shift the perceived depth of the scene to a range, which is comfortable for the user, based on the employed display [11]. Our hardware and also most other consumer electronics, enable such an algorithm by default. The algorithm interferes with our camera calibration and the effects on depth perception are neither quantifiable, nor correctable, so we disabled it completely by setting the parameter auto-convergence to mode-disable. This is also necessary to ensure that the displayable depth range is not skewed and lies completely behind the display from the user perspective.

The rendering component takes two inputs, the full video frame and the marker homography \(H_M\) for each detected marker in the left image. The output is a side-by-side image with the two perspectives adjacent to each other. First the video frame is converted on the GPU and applied to a stereo plane in full width. Next, the virtual objects are transformed according to \(H_M\) and the scene is rendered to the left half. After applying the extrinsic matrix \(M_E\), the scene is rendered again on the right half. When the parallax-barrier is enabled, the side-by-side image is automatically converted by the display driver. With the described pipeline, stereoscopic AR can be switched on and off instantaneously, to enable direct comparison. In stereo mode, we achieved interactive frame rates for tracking and rendering with 25-30fps and 20-25fps respectively.

During the development we found some perception problems resulting from the employed display technology. The apparent accommodation-convergence-conflict, typical for autostereoscopic displays, decreased after a short period of use. What remained was the ghosting, which occurs when dark regions in one image coincide with bright pixels in the other image. Despite these shortcomings, initial users of the prototype reported an intense feeling of immersion and depth.

5 Experimental Evaluation

In order to evaluate its effect on the perception of spatial relation and to compare binocular disparity to other depth cues in handheld AR we conducted a user study to quantify a) accuracy and b) completion time of depth positioning tasks between real and virtual

---

\(^1\)http://www.nintendo.com/3ds
\(^2\)https://code.google.com/p/andar/
\(^3\)https://launchpad.net/artoolkitplus
\(^4\)https://github.com/MasDennis/Rajawali
real object
(b)
(c)
(d)
(e)

Tests were conducted in our laboratory with static illumination from the markers. The space was limited in the back by a wall. All large white rectangle to eliminate additional perceptual clues from used a marker board with six markers, which was overlayed with a native movement of the object. To improve tracking stability, we gesture on the touchscreen of the phone, which resulted in a relative line in the schematic. This was done using a virtual object could be varied parallel to the y-axis along the case, a pink cuboid standing on the right of the desk. The position position) of the virtual object with a real reference object, in our personal space in front of the user. A schematic overview of the distance.

We designed the task to represent our target area, limited to the same (a) shadow (b) texture (c) banknote (d) sphere (e) sphere

The most common experiments to study depth perception are verbal report, perceptual matching, and open-loop action-based tasks [17]. They are based on egocentric depth judgments performed by the test subject, which are either reported verbally, matched to a reference distance or recreated after the observation. In this study we choose perceptual matching, which is an action-based closed-loop task. The reasons for this are twofold: First, these tasks resemble several interactions found in hand-held AR. Second, due to the mismatch between the stereo-base of our prototype device compared to the human inter-ocular distance, we expect that first-time users are not able to correctly translate the perceived depth to a real world distance.

We designed the task to represent our target area, limited to the personal space in front of the user. A schematic overview of the study setup can be found in Figure 3. We displayed different objects, represented by the green cube on the left, floating above the flat surface on an office desk. The task was to match the depth (y-position) of the virtual object with a real reference object, in our case, a pink cuboid standing on the right of the desk. The position of the virtual object could be varied parallel to the y-axis along the teal line in the schematic. This was done using a vertical swipe gesture on the touchscreen of the phone, which resulted in a relative movement of the object. To improve tracking stability, we used a marker board with six markers, which was overlayed with a large white rectangle to eliminate additional perceptual clues from the markers. The space was limited in the back by a wall. All tests were conducted in our laboratory with static illumination from artificial light sources. Figure 2 gives an impression of the scene displayed on the mobile device.

We wanted to compare the influence of disparity to other depth cues typically found in AR scenarios. Therefore we devised five different scenes with distinct virtual objects, featuring a variety of depth cues (see Figure 2). The virtual cuboid in the first test case had the same dimensions as the real one and was always placed directly on the surface. This provided the visual height and the retinal size as additional cues. The two cubes were enhanced with drop shadow and texture respectively. The banknote is an object of known size. The sphere added no depth information. The tests same, shadow and texture also provided linear perspective as an auxiliary depth cue. All objects, except the banknote, had the same dimension along the y-axis as the reference object. We saw no need to restrict the motion of the participants for our experiment, allowing them to use motion parallax as another depth cue. They were instructed to stay behind the red line, marked with tape on the desk surface to hinder them from switching to a side or top-down view. Only few participants made extensive use of motion parallax during the experiment.

The main subject of our evaluation was the effect of stereopsis on hand-held augmented reality so each participant performed each test case with and without the stereo condition. Each combination of test case and display condition was repeated four times in a within-subject study with repetition, resulting in $5 \times 2 \times 4 = 40$ individual tests per subject. To counteract temporal effects, the test order was completely randomized for each participant. In addition, the starting position of the virtual object was selected randomly from ten different heights (evenly spaced between 50 and 140mm) and ten different depths (evenly spaced between 30 and 300mm), to reduce learning. This also ensured that the virtual object was placed in front as well as behind the reference object situated at 200mm.

Each of the 40 test conditions was started and ended with the press of a physical button located on the side of the device. Since the device was operated in landscape orientation, the button could be operated very similar to a shutter release known from a camera.

We recorded the final depth offset between virtual and real object and the task completion time as dependent variables. The latter is corrected to account for the fact that the visibility of the augmentation was lost for short periods when the tracking failed. An overview of the experiment variables can be found in table 1.

5.2 Participants

We recruited 24 participants but had to exclude two of them because of stereo blindness and another two because of technical problems during the experiment. The remaining 20 subjects (four female, age between 19 and 33, mean of 26) had normal or corrected-to-normal vision. Most of them were members of our faculty (students and employees). Although all of them had a technical background, only one participant reported to have had prior experience with autostereoscopic displays.
To monitor possible side-effects, every person started by filling out a Simulator Sickness Questionnaire (SSQ) [5]. The device was handed to the participant in a demonstration mode and they were informed about the display, the optimal viewing angle and distance. After that, they were given some time to familiarize themselves with the hardware, followed by a description and a dry run of the task with a simple cube. All of the 40 variations were tested in different test cases and display conditions. The error for the first test case, mean error is reduced by 31% and in the latter it is more precise. The test case sphere shows a strong overestimation of the distance to the reference object (M: 30.0mm, SD: 63.4mm) in stereoscopic, but not in the monoscopic case (M: 2.2mm, SD: 8.7mm). For all of the test cases the standard deviation in stereoscopic mode is smaller than the monoscopic one, which could already be an indication that it is more precise.

5.3 Results

Figure 4 provides an overview on the acquired data points for the relative depth offset between both objects with a boxplot for each test case and for both display variants. The center line at 0mm would be a perfect alignment, while negative values translate to the virtual object being closer to the viewer, positive values were further away.

The first observation is that test conditions same and shadow were positioned very accurately by all participants. For these two tests, there is almost no difference between the two display conditions with a Mean (M) of 3.7mm (Standard Deviation (SD): 9.0mm) in the monoscopic and 3.1mm (SD: 8.1mm) in the stereoscopic condition. And the same is true for test condition shadow at 4.6mm (SD: 17.8mm) without and 4.6mm (SD: 10.2mm) with stereoscopy. The positioning error of all samples in these two test cases, lies in the range of ±5% of the object distance, except for some outliers.

The means for each condition in Figure 5 together with the standard error. The results indicate that there is a significant main effect of stereoscopy on the depth error. To investigate the effect of the different conditions on the depth error, we performed a 2 × 5 repeated measure ANOVA based on the absolute error. The results indicate that there is a significant main effect of stereoscopy on the depth error ($F_{1,19} = 43.424, p < .001$). Upon further analysis of the individual test cases we could only find a significant improvement for the test conditions banknote ($F_{1,19} = 7.935, p < .05$) and sphere ($F_{1,19} = 12.114, p < .05$). In the first case, mean error is reduced by 31% and in the latter it is still 10% more accurate.

The effects in the other tests were not significant: same ($F_{1,19} = 0.221, ns$), texture ($F_{1,19} = 0.173, ns$) and shadow ($F_{1,19} = 0.016, ns$).

We computed the absolute depth positioning error and plotted the means for each condition in Figure 5 together with the standard error.
test conditions same and shadow were finished more quickly. In direct comparison, the effect of stereoscopic vision is either negligible or tends to increase completion time. Further analysis can only confirm a significant effect for the stereo condition in test case shadow ($F_{1,19} = 53.167, p < .01$), where completion time gets worse. During the experiment we could observe that the tasks were usually completed in several phases: During the first phase, the situation was assessed and the initial position estimated. This was followed by a rough positioning and several smaller refinements.

The results of the SSQ showed an increase in eye strain for 45%, blurred vision for 35%, dizziness with closed eyes for 25% and difficulties to focus for 20% of the subjects. There were also individual reports of degradation for the categories dizziness with open eyes, general discomfort, fatigue and difficulties to concentrate. All stated changes were only one step from none to slight or from slight to moderate.

After the individual sessions of the experiment, several participants spontaneously expressed their opinion regarding the display. Four of them preferred the stereoscopic mode, while seven other subjects would choose the monoscopic option. Some others declared to perceive no difference between the two. There were several reports of blurred or low resolution images, ghosting, flickering and low brightness, all of which can be traced back to the parallax-barrier display.

6 DISCUSSION

In this study, we compared accuracy and speed of depth perception for monoscopic and stereoscopic hand-held AR. Five different virtual objects had to be aligned with a real object in individual test cases. Each of them provided other monoscopic depth cues.

Almost all participants were able to complete test case same and case shadow with minimal positioning error, regardless of the display condition. Apparently, the monoscopic cues were so strong, that the binocular disparity provided no significant improvement. In retrospect, both test cases were very easy to adjust because of the clear distinction created by the virtual object standing (or being projected) on the table. With this information it was no problem to align them with the same edge on the real object without the need for other depth cues. Test case shadow still produced significant results for the completion time, showing that the stereoscopic condition took longer to align. One possible reason could be that the subjects needed more time to adjust to the autostereoscopic display. This could also be an indication for confusion resulting from contradicting depth cues.

The test cases sphere and banknote did show significant improvements of positioning accuracy in the stereoscopic condition. It should be noted that these objects did not provide any cues based on linear perspective. Out of the nine participants who ranked the difficulty of the tasks, four named the sphere, three the banknote and another two voted for the textured cube. Test case texture did not show any significant effect of stereoscopy on positioning error. This is similar to the findings of Kerber et al. [6], who investigated minimal depth discrimination distances between virtual cubes using the same hardware. They argue that object size already provides a strong depth cue which is conflicting with binocular disparity, but this cannot fully explain why the effect of the stereo condition is stronger in case of the sphere than in texture condition of our study.

In conclusion, our study shows that the use of autostereoscopic displays can significantly improve depth perception in hand-held AR. This is especially the case when the scene offers no other dominant depth cues. These results go in line with the weak fusion model of depth perception [10], in which all depth cues contribute to depth perception with different weights. However, binocular disparity seems to have smaller impact than expected in this scenario.

One of the major drawbacks in terms of user experience was the display technology. The SSQ also showed that several of the participants were affected by the test procedure, although the impact was minor and could also result from the high level of concentration. We think that diminished resolution and brightness, as well as ghosting could be reduced in future hardware generations, but the accommodation-convergence mismatch is not as easy to overcome.

One effect we witnessed during our study, was that the participant, who declared to use a hand-held autostereoscopic game console regularly, performed better than average and also did not show any signs of visual fatigue. We speculate that this could be a sign of adaption to the technology, but this needs to be evaluated in a long term study, which was not done to date.

7 CONCLUSION AND FUTURE WORK

In this paper we presented an approach to improve depth perception in hand-held augmented reality using autostereoscopic displays. We demonstrated the feasibility to implement such a system on an off-the-shelf smartphone. The relevance of binocular depth cues were studied in an experiment conducted with our working prototype. The results indicate that stereopsis significantly lowers positioning error between real an virtual objects in challenging compositions.

It should be noted that in some of our test cases other indicators for alignment had a much larger influence on the performance. Therefore it seems reasonable to suggest to designers of AR application, to include different forms of depth indicators. Example for these are auxiliary augmentations [8], depth blur [14] or artificial holes to overcome occlusion [15]. While coinciding cues should reinforce each other, contradictions are not as easily resolved. Future studies need to examine how hand-held stereoscopic AR can be used best in combination with other techniques and in more complex environments.

ACKNOWLEDGEMENTS

This work was supported in part by the German Federal Ministry of Education and Research (BMBF).

REFERENCES


