Citation for final published version:

Kinateder, Max, Gualtieri, Justin, Dunn, Matt J. , Jarosz, Wojciech, Yang, Xing-Dong and Cooper, Emily A. 2018. Using an augmented reality device as a distance-based vision aid - promise and limitations. Optometry and Vision Science 95 (9) , pp. 727-737. 10.1097/OPX.0000000000001232

Publishers page: http://dx.doi.org/10.1097/OPX.0000000000001232

Using an Augmented Reality Device as a Distance-Based Vision Aid – Promise and Limitations

Max Kinateder Ph.D.[1*], Justin Gualtieri[2], Matt J. Dunn Ph.D. MCOptom[3], Wojciech Jarosz Ph.D.[2], Xing-Dong Yang Ph.D.[2], & Emily A. Cooper Ph.D.[1,2]

[1]Department of Psychological & Brain Sciences, Dartmouth College, Hanover NH 03755
[2]Department of Computer Science, Dartmouth College, Hanover NH 03755
[3]School of Optometry and Vision Sciences, Cardiff University, Cardiff UK CF24 4HQ

[*] Corresponding author:
email: max.kinateder@dartmouth.edu
6207 Moore Hall, Dartmouth College, Hanover NH 03755, USA

Word count: 7670
Number of tables: 2
Number of figures: 7

Abstract

Significance: For people with limited vision, wearable displays hold the potential to digitally enhance visual function. As these display technologies advance, it is important to understand their promise and limitations as vision aids.

Purpose: Test the potential of a consumer augmented reality (AR) device for improving the functional vision of people with near-complete vision loss.

Methods: An AR application that translates spatial information into high contrast visual patterns was developed. Two experiments assessed the efficacy of the application to improve vision: an exploratory study with four visually impaired participants, and a main controlled study with participants with simulated vision loss (N = 48). In both studies, performance was tested on a range of visual tasks (identifying the location, pose and gesture of a person, identifying objects, and moving around in an unfamiliar space). Participants' accuracy and confidence were compared on these tasks with and without augmented vision, as well as their subjective responses about ease of mobility.

Results: In the main study, the AR application was associated with substantially improved accuracy and confidence in object recognition (all Ps < 0.001) and to a lesser degree in gesture recognition ($P < 0.05$). There was no significant change in performance on identifying body poses, or in subjective assessments of mobility, as compared to a control group.

Conclusions: Consumer AR devices may soon be able to support applications that improve the functional vision of users for some tasks. In our study, both artificially impaired participants and participants with near-complete vision loss performed tasks that they could not do without the AR system. Current limitations in system performance and form factor, as well as the risk of over-confidence, will need to be overcome.

Keywords: augmented reality; assistive devices; low vision and blindness

For the millions of people who are affected by low vision and blindness, independence and mobility pose daily challenges.[1-3] To address these challenges and improve the functional vision of this population, a range of assistive tools have been developed, including vision aids and sensory substitution devices. Recently, available tools have included custom head-mounted display (HMD) systems designed to digitally enhance visual information, such as Jordy (Enhanced Vision, Huntington Beach, CA), LVES,[4] eSight (eSight, Toronto, ON), and NuEyes (NuEyes, Newport Beach, CA). The basic principle of these HMDs is to substitute the image cast by the world on the retina with an enhanced view. Outward-facing cameras capture live video of the world in front of the user; this video is processed to increase visibility via magnification or contrast enhancement, and then shown in (near) real-time to the user through a pair of micro-displays positioned in front of the eyes.[5-7] This is called a ‘video see-through display’ because although the system is mobile, the users’ eyes are covered by opaque screens. While these systems are promising and can measurably increase functional vision,[6] they also tend to suffer from temporal lag, cumbersome hardware, and reduced visual field. To date, no video see-through system has been widely adopted.

At the same time, HMDs have emerged as a popular platform for mass consumer electronics, with a range of companies selling these systems to general consumers for virtual and augmented reality (VR/AR) applications. In particular, *optical* see-through AR systems – such as Glass (Google, Mountain View, CA) and HoloLens (Microsoft, Redmond, WA) – can augment vision without having to cover the eyes with an opaque screen. These commercial products also benefit from the cost-savings of mass production, improvements in form factor, and the ability to flexibly support a range of software applications (“apps”). Despite the lower contrast typical of see-through displays, these AR systems have several potential advantages compared to *video* see-through displays. For example, the user’s natural field of view is intact, and their eyes are un-occluded. Thus, the incorporation of assistive features into a consumer AR system provides a potential new avenue for broadening the impact of this technology on the low vision and blind population, much like consumer smartphones have broadened the availability of hand-held assistive tools.[8]

One early study used off-the-shelf HMDs to build a see-through visual multiplexing device for visual field loss,[9] but at the time additional custom hardware was required to achieve the desired effect. A more recent study examined visual acuity and sensitivity for text and shapes presented on a see-through AR system, showing that a variety of virtual content can be visible to visually impaired users on a consumer system.[10] However, no specific assistive applications were explored. Another recent study showed that overlaying enhanced edge information on a see-through HMD can increase contrast sensitivity in simulated visual impairment.[11] Here, we build on this prior work to examine alternative avenues for visual enhancement using consumer AR.

The question how best to augment visual information is still an open one.[5, 12-14] Particularly in complex natural environments, overall edge or contrast enhancement may not make individual objects and elements easier to perceive for individuals with near-complete vision loss (i.e., individuals with severely impaired vision or legal blindness).[15] Instead, selectively enhancing only the edges that indicate object boundaries may simplify complex visual patterns so as to help people with severely impaired vision parse natural scenes.[16-20] In particular, a few previous studies have employed a 'distance-based' enhancement system that translates the distance of points in front of the user into pixel brightness values, and showed that visually impaired users wearing this video system could perform a visual search task while seated,[16] and collided with fewer obstacles in a mobility task.[21] A similar approach was recently implemented in a custom-built see-through system.[22] Here, we examine the ability of emerging consumer AR hardware (**Figure 1**) to implement a similar distance-based visual augmentation, with a focus on usability for individuals with near-complete vision loss. We focus on this group specifically because prior work and our own pilot testing suggest that they may be the most likely to find utility in distance-based information. Thus, we test the hypothesis that distance-based AR can improve functional vision in this target population for a range of tasks. We develop an application to run on the HoloLens that translates spatial information from the physical environment into an AR view containing simplified patterns with high-contrast edges between objects at different distances. We then examine the impact of the application on performance of a range of visual tasks in an exploratory study with visually impaired users (*N* = 4) with a range of etiologies, and in a main study using a larger sample (*N* = 48) of people with simulated visual impairment. We focus on understanding existing strengths, areas of potential, and current limitations.

Figure 1 about here.

Methods

Hardware

The HoloLens is a head-mounted AR system that can display 3D virtual surfaces within the physical environment.[23] The system includes two see-through displays that subtend approximately 30° horizontally and 17.5° vertically in each eye (**Figure 1A**, red arrows). A set of sensors (**Figure 1A**, blue dashed box) – including four scene-tracking cameras, an infrared-based depth sensor, and an inertial measurement unit – continuously track the user's position and orientation in the environment. As the user moves around, the HoloLens also measures and stores the dimensions and shape of the physical space around them, creating a 3D reconstruction of the surrounding environment. This 3D reconstruction is provided to developers as a triangle mesh, in which the number of individual triangles used to define the environment per unit area determines the resolution and detail of the 3D map. User input is accepted via multiple channels, including speech, hand gestures, and Bluetooth devices. All computation is completed on board, so the system is untethered (**Figure 1B**) and has a battery life of 2-3 hours with active use. It weighs approximately 580 grams.

Application development

Software development was performed using Microsoft's HoloToolkit and Unity (Unity Technologies, San Francisco CA, USA). We developed an application that measures the distance of surfaces and objects in the environment from the user by accessing the user's position and the 3D environment map. The application discretizes these distances into a set of bands, each with a unique color and intensity value. The bands are directly overlaid semi-transparently on the environment in stereoscopic 3D when viewed through the displays (**Figure 2A,B**), creating an AR environment that is a mixture of real and virtual surfaces. The AR environment has a simplified visual geometry, with high contrast-edges between objects and surfaces at different distances from the observer, which we hypothesize is more easily interpretable by people with impaired vision relative to the original view.[16, 17, 19] When using the system, the natural field of view is unrestricted, so the appearance is similar to having a window into the AR environment through the HoloLens display (see above). As the user moves around the environment, the colors change to reflect the distances from the current viewpoint. The mapping between distance and color is arbitrary. We created 18 unique mappings to enable customization for different levels of visual impairment and color vision (**Figure 2C** shows 9 examples). Some mappings transition between two colors from high to low saturation (left column); some transition from white to one color (middle column) and some transition from high to low opacity (right column). Because the HoloLens displays can produce light but cannot occlude it, transitions from white to black are not possible. In addition, the overall luminance and opacity of the overlays is adjustable, which is useful for cases in which a user is particularly light sensitive, or for transitioning between environments with differing ambient light levels. The source code for our application is freely available for research purposes.

Figure 2 about here.

In Experiment 1, we allowed users to select any one of the 18 mappings that created the most visible contrast between the foreground and background of a scene. In Experiment 2, we used two different mappings (red-to-blue, shown in **Figure 2B**, and high-to-low opacity). In both experiments, the update rate for the display and motion tracking was set to 60 Hz, and the resolution of the 3D environment mesh was set to the highest density that produced noticeable improvements in 3D detail (~2000 triangles per cubic meter). There was a one second delay between subsequent mesh updates, which was necessary for the system to scan and process the updated mesh. Thus, all visual identification tasks were performed with the target person, object, or gesture held stationary. Due to the fast tracking of user-generated motion, there was no noticeable lag associated with body or head movements.

The number of discrete color bands was set to 10 and distances closer than 0.5 m were not augmented, so as not to impede near-vision. In Experiment 1, the first band covered 0.5 m to 1.5 m, the eight middle bands were each 0.25 m wide, and the final band covered distances beyond 3.5 m. In Experiment 2, the first band extended to only 1.0 m, and all other bands were also moved closer by 0.5 m accordingly.

Experiments

All participants in both experiments gave written informed consent and were compensated. The procedures were approved by the Dartmouth College Institutional Review Board and comply with the Declaration of Helsinki. The procedures and main hypotheses of Experiment 2 were preregistered on AsPredicted.org (#2870). For clarity, **Table 1** provides a summary of the participants, tasks, and number of trials conducted in each experiment.

Table 1 about here.

Experiment 1: Exploratory study with visually impaired participants

Participants

Four participants were recruited via an email advertisement. **Table 2** provides individual information about each participant. Note that Participant 4 works as a professional accessibility services manager. Participants were recruited with a range of conditions causing generalized vision loss and in some cases, visual field restriction.

Table 2 about here.

Customization

The experimenter calibrated the HoloLens for each participant in a two-step procedure. First, all pixels were turned on uniformly, and the device was adjusted to make sure that the displays were visible and the overall brightness was at a comfortable level. Next, the experimenter stood 1.5 m from the participant and turned on an initial color setting. The participant looked around and determined whether they could visually identify the location and shape of the experimenter's body. At this stage, each participant indicated that they could see the experimenter. We then interactively determined the color setting that created the strongest perceived contrast between foreground and background. Finally, the experimenter stepped slowly backwards to confirm that the visible contrast changed with distance. While this approach limited our ability to combine results across participants, due to the range of visual system pathologies present, it maximized the potential impact for each individual. Of the four participants, one selected red-to-blue (Participant 1), two selected yellow-to-blue (Participants 2 and 3), and one selected white-to-blue (Participant 4).

Tasks

We conducted four naturalistic tasks, each consisting of two blocks of five trials. The first block was completed with the AR turned off (baseline) and the second block with the AR on, and the trial order within each block were pseudo-randomized. Participants' performance (correct/incorrect) and confidence (from 1 "it's a guess" to 3 "very certain") were recorded for each trial. Tasks were selected to represent different levels of difficulty in visual identification, as well as mobility. Participants performed all tasks in an indoor space with typical overhead lighting. Prior to starting each task, participants performed a brief practice both with and without visual augmentation.

Person localization: Participants sat in a chair and a life-size cutout figure of a person was placed 1.8 m away from them. The location of the figure was pseudo-randomly assigned to one of five positions (-45.0°, -22.5°, 0.0°, 22.5°, 45.0° from 'straight ahead'). On each trial in the baseline and AR blocks, the participants indicated the location of the figure using a laser pointer. The experimenter scored hits (1), near misses (0.5, the laser pointer missed the cut-out figure only slightly) and misses (0). After each trial, participants rated their confidence.

Pose recognition: On each trial, the experimenter stood 1.5 m from participants and held their arms in one of five different poses (arms straight out to the side, arms up forming a "Y," arms above the head forming an "O," one arm straight up / one arm straight down, one arm bent down at elbow / one arm bent up at elbow). The experimenter wore a black long-sleeve jacket and the wall behind them was beige with some decorations, so that there was high contrast between the foreground and background even without any augmentation. Participants mirrored each pose with their arms and indicated their confidence. The response was recorded with a photograph and later scored by a naïve judge on a 3-point scale with 0 indicating incorrect, 0.5 partially correct, and 1 fully correct.

Object recognition: Participants identified objects that were placed one at a time on a table 1.5 m in front of them, and reported their confidence. The objects were a spray bottle, table lamp, square wicker basket, recycling bin, and fake plant (**Figure 3A**). Prior to starting the task, participants were given time to touch and look at each of the objects and identify them verbally.

To control for memory effects, the experimenter read aloud the list of objects before each block. Participants responses were scored as either incorrect or correct.

Mobility: Participants walked forward from a fixed location and stopped when they identified an obstacle in their path (a white portable room divider 1.7 × 1.6 m). All participants except Participant 4 completed the task without a cane. In each trial, the obstacle was placed at a pseudo-randomly selected location between 5.5 m and 7.5 m from the starting position. After participants stopped, the experimenter measured the distance between them and the obstacle using a laser range finder. Confidence scores were not collected, because participants were instructed to stop as soon as they detected the obstacle.

Experiment 2: Controlled experiment with simulated vision loss

Sample

Forty-eight participants (mean age: 21.15, 34 female) were recruited, all with normal or corrected-to-normal visual acuity (0.00 logMAR or better) and normal stereoacuity (70 arcsec or better) assessed with a Randot Stereo Test (Precision Vision, LaSalle, IL, USA). During all tasks, participants wore a pair of swim-goggles modified binocularly with Bangerter occlusion foils (type LP; Ryser Optik, St. Gallen, Switzerland), which degrade visual acuity uniformly across the visual field[24]. The LP-type foils simulate visual acuity at the level of perceiving hand movements, with some rough shapes and forms distinguishable under typical indoor lighting. For each participant, we verified that the simulators resulted in letter acuity less than 1.60 logMAR (approx. 20/800), inability to count fingers at 1.0 m, and intact perception of hand movements. One session was repeated due to technical errors.

Conditions

Participants were randomly assigned to one of three groups (n = 16). In the *color group*, the red-to-blue AR color mapping was used (**Figure 2B**). In the *opacity group*, the bands had differing levels of opacity: near distances were most opaque and distances beyond the 9th band were fully transparent. In the *control group,* participants were told that the HoloLens would augment their vision, however, no actual augmentation was displayed (at the start of each task for which vision was supposed to be augmented, the HoloLens screen flashed blue and faded back to being fully transparent). This group was included to examine potential practice effects or increases in effort/attention associated with the knowledge of augmented vision.

Visual identification tasks

Participants performed three identification tasks, each consisting of two blocks of six trials (the first block with the AR turned off and the second block with the AR on). The overall procedure used was the same as the exploratory study, but the study was carried out in a different location and with some differences in the tasks. Three naïve judges scored pose and gesture recognition accuracy and their ratings were averaged to determine the final score.

Pose & object recognition: These tasks were performed in the same manner as described in Experiment 1, with the exception that the viewing distance for poses was 2.2 m. A sixth pose ("both arms straight up") and object (stack of books) were also included. The inter-rater reliability of the scoring for poses was 0.78 (Fleiss Kappa), suggesting substantial agreement (defined as 0.61-0.80).[25]

Gesture recognition: To assess the spatial resolution of the augmented vision, the experimenter stood 1.2 m from the participants and made one of six gestures with their right hand held to their side (thumb-up, shaka ["hang loose"], open palm, fist, peace sign, okay). The participants mirrored the hand gesture and indicated their confidence. Responses were scored as for the poses and inter-rater reliability was 0.63.

Figure 3 about here.

Mobility task

Participants explored a room (5.3 m × 3.6 m) with an unknown layout in three trials. On each trial, the furniture in the room was arranged in one of three different layouts (selected pseudo-randomly) and the participants were given 60 sec to complete the task (**Figure 3B**). On the first trial, the AR remained off (baseline). There were two test trials: one in which the AR was on, and another in which a white cane was used as an assistive tool. The ordering of these two trials was determined pseudo-randomly. Prior to the cane trial, participants practiced using the cane in a different room. After each trial, participants rated their level of agreement on a scale of 1 (strongly disagree) to 7 (strongly agree) with four statements: “Overall, I felt comfortable while exploring the room”, “I felt unlikely to run into things”, “It was easy to navigate the space”, and “I felt that my vision provided useful information”. After all trials, participants indicated whether baseline, AR, or cane was the best with respect to each of these statements. Since we used the same room with different layouts, the HoloLens’ storage of overlapping spatial meshes could cause technical issues. Thus, between trials we cleared the system memory and circled the room once to orient the system to the new layout (note that this problem does not occur if the system is moved to a new room).

Data analysis

All data were analyzed using the R Environment for Statistical Computing, version 3.3.2.[26] For Experiment 1, in some cases participants were unable to detect any visual information during the baseline trials and did not provide guesses. On these trials, confidence was scored as zero (note that this was the case for all baseline trials for Participant 4). For Experiment 2, effects of the independent variables (*experiment group* [control/color/opacity] between subjects, and *trial block* [baseline/AR] within subjects) were assessed using repeated measures ANOVAs (significance level of $P < 0.05$). For post-hoc analyses, p-values were Bonferroni corrected. Normality of data from Experiment 2 were tested using Shapiro-Wilk tests. For gesture and pose recognition in Experiment 2, analyses were performed on the average accuracy ratings of the three judges. Due to technical errors, data from one trial in Experiment 1 and one trial in Experiment 2 were not recorded. The raw response data and analysis code are provided on publicly accessible repositories.

Results

Experiment 1

Accuracy and average confidence ratings for each of the four participants in the visual identification tasks are shown in **Figure 4A-C**. Each pair of colored bars shows the results for an individual participant's baseline and AR trials. Participants 1, 2 and 3 were able to complete the person localization task consistently both with and without AR, and reported high confidence (**Figure 4A**). Participant 3 (brown bars) indicated after the task that the augmentation made her more confident (despite her ratings being similar). However, Participant 1 (magenta bars) remarked that the checkered shirt of one experimenter was actually more visible without the augmentation. Participant 4 (yellow bars) was unable to locate the figure without AR, but correctly located it on 80% of trials with AR, with medium confidence. Similar patterns were observed for pose recognition (**Figure 4B)**. Participants 1 and 2 performed the task with high accuracy and confidence, but for this task Participant 3 had lower accuracy overall (compared to person localization), and reported higher confidence with AR. Participant 4 was unable to perform the task at baseline, but obtained reasonably accurate performance (with low confidence) with AR. Qualitatively, all but Participant 2 improved in object recognition in the AR block (**Figure 4C)**, while Participant 2 decreased slightly both in performance and confidence.

Figure 4 about here.

The results for the mobility task are shown in **Figure 4D**, in terms of the average distance each participant required to detect the obstacle and stop walking. In most trials without AR, Participants 1, 2, and 3 detected the obstacle one or two steps before reaching it. Participant 1 detected the obstacle on average at a similar distance in the baseline and AR blocks (1.4 and 1.6 m). However, he reported using a different strategy in the two conditions: in the baseline trials, he used the contrast between the obstacle and the background, when using AR, he instead relied on the color-distance information. This participant also indicated that the augmentation worked well for him to identify walls, and used it to guide himself to stop each time he returned to the starting position. Participants 2 and 3 both tended to detect the obstacle in the AR block from approximately 3 m, which roughly matches the onset distance of the farthest color transition; however, Participant 3 indicated that using a cane would be simpler. Participant 2 walked fastest, and on some trials experienced issues with the color map not updating quickly enough. Participant 4 could not detect the obstacle visually at baseline, so he used his cane. In one baseline trial and one AR trial, the participant changed direction prior to reaching the obstacle and thus never located it. However, on each of the AR trials, he detected the obstacle visually before hitting it with his cane, with an average distance of 1.88 m.

Other responses

Participants also reported on the strengths and weaknesses of the application and the hardware after completing all tasks. Participant 1 stated that if the hardware had the same form factor as a pair of glasses, it would be useful, and that providing distance information relative to the head was preferable for him than receiving this feedback on other parts of the body (like the arm). Overall, he said the system was somewhere between distracting and helpful. Participant 2 stated that overall his vision was worse with the overlays, and that the lag time was a problem (as we observed during the mobility task). Participant 3 also expressed that the current form factor of the system was undesirable, but that she might find the system particularly helpful at night.

Participant 4, whose vision was more strongly impaired than the others' and most improved when using the AR system, noted that he had to move his head around more in the identification tasks. This may reflect the limited display size in the visual field. However, unlike the other participants he indicated that the device was comfortable as is and that the form-factor was not an issue.

Overall, these results suggest that improvements in functional vision (particularly for object identification and obstacle detection during mobility) may be achievable with the AR system, but indicate that the usefulness of the distance-based augmentation likely varies by task and visual ability. In addition, these results on their own do not rule out the possibility that any objective or subjective changes in vision could be due to increased attention, effort, or practice during the trials with augmented vision, due to the novelty of using AR.

Experiment 2

In this main study, we examined the potential changes in functional vision created by the AR system in a larger sample of participants with simulated near-complete vision loss. We also examined the potential impact of the system novelty on our measures of performance by inclusion of a control group.

Visual identification tasks

The results from each of the three visual identification tasks for mean accuracy (top row) and confidence (bottom row) are shown in **Figure 5**, separately for the *control group* (gray bars), *color group* and *opacity group* (orange bars). Recall that the procedure for the *baseline* blocks (light shaded bars) was identical for each group, so variability across groups can be attributed to random variance, and that the *control group* was told they would have augmented vision, but after a brief flash the HoloLens display was actually turned off. The three tasks were selected to range from easy (pose recognition) to difficult (object and gesture recognition) when performed at baseline. This is reflected by the fact that baseline accuracy and confidence are overall high for pose recognition and relatively low for object and gesture recognition. A useful vision aid should ideally improve performance on tasks that are challenging, but importantly, it should also not degrade performance on tasks that are already easily accomplished with un-augmented vision.

Figure 5 about here.

First, we examine the effect of the augmentation on accuracy on each task. For pose recognition (**Figure 5A**), there were no significant main effects or interaction terms for experimental group (*control*/*color*/*opacity*) or trial block (*baseline*/*AR*) variables (*experiment group*: $F(2,45) = 0.72$, $P = 0.49$, $\eta_p^2 = 0.03$; *trial block*: $F(1,45) = 0.82$, $P = 0.37$, $\eta_p^2 = 0.02$; *interaction*: $F(2,45) = 0.15$, $P = 0.86$, $\eta_p^2 = 0.01$). Thus, while performance did not significantly improve with AR on this task, it also did not get worse. This is not entirely surprising, because performance was already quite high at baseline due to the high visual contrast (average percent correct across all groups was 61.3%). For object recognition (**Figure 5B**), significant main effects of experiment group and trial block were mediated by a significant interaction term ($F(2,45) = 13.01$, $P < 0.001$, $\eta_p^2 = 0.37$). Post-hoc comparisons showed that performance improved significantly during the AR block in the *color* and *opacity* groups, but not in the *control* group (*control:* $t(95) = 0.18$, $P = 0.86$, d = 0.03; *color:* $t(95) = 7.59$, $P_{corrected} < 0.001$, d = 1.10; *opacity:* $t(95 = 7.36$, $P_{corrected} < 0.001$, d = 1.06). Similarly, there was a significant interaction term ($F(2,45) = 3.66$, $P < 0.05$, $\eta_p^2 = 0.14$) in the gesture recognition task (**Figure 5C**), reflecting the fact that participants in the *opacity* group performed better in the AR block ($t(95) = 2.88$, $P_{corrected} < 0.05$, d = 0.42). This suggests that participants were able to use the information provided by the augmented vision to more accurately perceive the shape of the objects and the form of a hand gesture. In the case of gestures, the improvement was minor and likely not of practical use. For objects however, this improvement was substantial: the average percent correct was 65.0% using AR as compared to 19.4% without (over six trials). This is a promising amount of improvement, particularly considering that the level of simulated visual impairment was so severe. The ability to reliably recognize everyday objects visually with this system thus represents a practical improvement in functional vision.

Similar to the accuracy results, confidence ratings showed that participants overall rated their confidence to be highest in the pose recognition task and lowest in the object and gesture tasks. The confidence ratings for poses are shown in **Figure 5D**. As with accuracy, there were no significant main effects or interaction terms (*experiment group:* $F(2,45) = 0.64$, $P = 0.53$, $\eta_p^2 = 0.03$; *trial block:* $F(1,45) = 1.32$, $P = 0.26$, $\eta_p^2 = 0.03$; *interaction:* $F(2,45) = 1.06$, $P = 0.35$, $\eta_p^2 = 0.05$). For object recognition (**Figure 5E**), however, significant main effects were again mediated by a significant interaction term ($F(2,45) = 5.55$, $P < 0.01$, $\eta_p^2= 0.20$). Participants reported higher confidence during the AR block in both the *color* and the *opacity* groups, but not in the *control* group (*control:* $t(95) = 2.01$, $P_{corrected} = 0.19$, d = 0.29; *color:* $t(95) = 6.75$, $P_{corrected} < 0.001$, d = 0.97; *opacity:* $t(95) = 8.30$, $P_{corrected} < 0.001$, d = 1.20). Finally, confidence ratings for gesture recognition (**Figure 4F**) also showed a significant interaction term ($F(2,45) = 3.48$, $P < 0.05$, $\eta_p^2= 0.13$), reflecting higher confidence in the AR block in the *color* and *opacity* groups (*color:* $t(95) = 4.94$, $P_{corrected} < 0.001$, d = 0.71; *opacity:* $t(95) = 5.39$, $P_{corrected} < 0.001$, d = 0.78).

These results show that participants tended to be more confident in the two more difficult tasks when using the AR system. This makes sense for object recognition, in which their performance improved with AR. The confidence that a user has with their augmented vision likely plays a key role in how willing they are to rely on visual information and perform tasks independently. It is interesting that confidence increased for gesture recognition as well, because performance was only modestly impacted. In the next section, we report an exploratory analysis assessing the possibility that using augmented vision might produce overconfidence: an increase in confidence even when perceptual judgments are incorrect. In this and subsequent analyses, we combine the two test groups (*color*/*opacity* are grouped together as *test*), because the pattern of results were highly similar.

Confidence as a function of performance

In all visual identification tasks, confidence ratings and performance were significantly positively correlated (poses: $r = 0.53$, $P < 0.001$; objects: $r = 0.43$, $P < 0.001$; gestures: $r = 0.17$, $P < 0.001$). **Figure 6** shows the average confidence ratings for each task separately for trials in which participants gave correct or incorrect responses (for pose and gesture recognition, trials with a score greater than 0.75 were categorized as "correct", trials with scores lower than 0.25 were categorized as "incorrect"). Across all tasks, experiment groups, and trial blocks, participants tended to report higher confidence in trials in which they gave correct answers. Interestingly, partially overlapping t-tests (Bonferroni corrected for 12 comparisons; note that the number of observations in each bin varied) revealed that participants in the *test* groups reported higher confidence in the AR block, even when they gave incorrect answers (orange bars).[27] The only exception is the incorrect trials for pose recognition (**Figure 6A**). This overconfidence was not observed in the *control* group (gray bars). This underscores the importance of considering how to provide feedback and training to help users understand how reliable their vision is when they use an unfamiliar assistive device.

Figure 6 about here.

Mobility task

Figure 7 shows the results from the participants' responses after the mobility task. Rather than detect a single obstacle, in this task participants were given time to freely explore an unfamiliar room. For simplicity, responses are plotted as difference scores by subtracting out each participant's response in the baseline trial. Overall, these results show that reported improvements were similar across both the *control* and *test* groups, suggesting that the subjective assessments used in this task did not measure any potential effects of the AR system on mobility. In both the *control* and *test* groups, participants tended to report feeling less likely to collide with obstacles when using a cane and when using AR (**Figure 7A**). An ANOVA revealed only a main effect of trial type (*trial type [baseline/cane/AR]:* $F(2,92) = 22.72$, $P < 0.001$, $\eta_p^2 = 0.33$; *experiment group [control/test]:* $F(1,46) = 0.02$, $P = 0.89$, $\eta_p^2 < 0.01$; *interaction:* $F(2,92) = 0.80$, $P = 0.45$, $\eta_p^2 = 0.02$), and differences relative to baseline were statistically significant for all conditions except when the control group used the cane (*test/AR*: $t(31) = 3.45$, $P_{corrected} < 0.01$; *test/cane*: $t(31) = 6.01$, $P_{corrected} < 0.001$; *control/AR*: $t(15) = 3.76$, $P_{corrected} < 0.01$; *control/cane*: $t(15) = 2.31$, $P_{corrected} = 0.14$). When comparing collision risk, 65.5% of the *test* group preferred the cane and 34.5% preferred AR. In the *control* group, 56% and 38% preferred the cane and AR, respectively. Similarly, participants in both groups tended to report that their vision was more useful with AR (**Figure 7B**). There was also a main effect of trial type on these responses (*trial type:* $F(2,92) = 13.14$, $P < 0.001$, $\eta_p^2 = 0.22$; *experiment group:* $F(1,46) = 0.10$, $P = 0.76$, $\eta_p^2 < 0.01$; *interaction:* $F(2,92) = 0.45$, $P = 0.64$, $\eta_p^2 = 0.01$), which reflected a statistically significant increase in both groups when using AR (*control:* $t(15) = 3.65$, $P_{corrected} < 0.01$; *test:* $t(31) = 0.10$, $P_{corrected} < 0.01$). When comparing usefulness of vision, 78.1% of the *test* group and 62.5% in the *control* group reported that AR was preferred. Because the *control* group experienced no real augmentation, these results together indicate that subjective ratings are likely an unreliable measure of mobility improvements in AR. For the two other statements (“Overall I felt comfortable while exploring the room.”, “It was easy to navigate the space.”), no significant effects of using a cane or AR were found.

Figure 7 about here.

Discussion

The advent of mass-market consumer AR systems, together with the rapid development of assistive mobile technology, holds substantial promise for visually impaired individuals. Although the diversification and increased availability of high-tech tools might assist and one day even eliminate the need for biological vision in performing many day-to-day tasks, the precise potential benefits and challenges are still unclear. Here, we present two experiments using an application developed and deployed on a consumer AR device, which provides high-contrast, customizable distance information overlaid in the user's field of view. The results suggest areas in which current AR systems may be used to improve functional vision, and where they fall short.

Overall, our findings support previous work that simplifying visual scenes can be helpful for people with severely impaired vision, and show that this approach can be implemented in a see-through HMD display.[16, 17, 19, 21] However, our studies indicate that the utility of the current system varies substantially as a function of task. Experiment 1 also suggests that this particular system may not be desirable in all forms of vision loss, both because visual detail from surface texture can be lost, and because the resolution of the HoloLens 3D spatial mesh is limited. This does not preclude the potential utility of AR for these users, who may instead benefit from overall edge or contrast enhancement.[11] The flexibility of consumer devices provides a potential platform to create a variety of applications from which a selection can be made depending on a user's level of visual ability. However, the type of applications that are possible, and how they should differ for different users, is an area that requires further research. Although low vision and blindness simulators are frequently employed to examine task performance in controlled settings,[11, 28, 29] future work should examine systematically how the acuity levels and visual field loss associated with specific etiologies may be addressed with AR.

Major limitations of the current HoloLens system include the fact that it only updates distance information at up to 1 Hz, so visual perception of fast moving objects may be degraded. However, the display can provide low-latency self-motion information because it builds up a stable 3D map as the user moves around a stationary environment. Nonetheless the lag and limited range of the mapping are clear limitations of the device, which will hopefully improve with the next generations of HMDs. As 3D sensing technologies improve, the ability to quickly update both self and environmental motion will be essential. At the same time, the portion of the visual field covered by the see-through display of the HoloLens is quite limited (30° horizontally). Key information for several activities, such as navigation, may often fall in the peripheral visual field, so improvements in the display size are highly desirable. In addition, the distance-based nature of the current system means that regions of high visual contrast but low depth variance would likely be degraded visually. Future generation systems could detect object boundaries using a combination of depth and image-based measures.[30] In this case, it may be possible to dynamically adjust the pattern or opacity of overlaid depth information to minimize interference with other visual details. Finally, in its current state, the display brightness is limited and best-suited for indoor environments.

Our results also suggest an interesting effect of AR on visual confidence. Visual confidence (i.e., an observers' ability to estimate the reliability of their own perception[31]), might be of particular importance for users who adopt HMD-based tools. While people have extensive experience with which to estimate the reliability of their unaided vision, they have no immediate access to quantitative diagnostics of an HMD. As with other assistive devices, training, practice, or calibration is likely to be necessary in order for users to learn the correct level of visual confidence. Here, we found that accuracy was indeed positively correlated with confidence. However, we also found that when participants used augmented vision, their visual confidence was higher compared to baseline, even when they gave *incorrect* answers. However, it is important to note that this observation was made from a sample of participants with simulated, temporary visual impairments, and thus may not generalize to other populations. Future research will therefore need to explore our understanding of visual confidence in AR.

Based on the results and feedback in these studies, several future directions are conceivable. For instance, recent advantages in computer vision could be harnessed to develop "smart" overlays that, for example, are able to highlight flat and uneven surfaces and identify stairs, apertures, or even people. In addition, more sophisticated algorithms to automatically provide simplified and enhanced spatial information could potentially be implemented in real-time

AR.[32, 33] The rapid developments in mobile electronic consumer devices' computing power together with universal platforms for application development, provide vast opportunities to implement and improve visual assistive technology.

Acknowledgements

The authors want to thank the Carroll Center for the Blind, Bruce Howell, and Robin Held for their advice and assistance; Klara Barbarossa and Jonathan Huang for help with data collection; and two anonymous reviewers for helpful feedback on the manuscript. The research was supported by gifts provided by Microsoft and Oculus.

References

1. Varma R, Vajaranant TS, Burkemper B, et al. Visual Impairment and Blindness in Adults in the United States: Demographic and Geographic Variations from 2015 to 2050. JAMA Ophthalmol 2016;134:802-9.

2. Massof RW. A Systems Model for Low Vision Rehabilitation. Ii. Measurement of Vision Disabilities. Optom Vis Sci 1998;75:349-73.

3. Stelmack JA, Szlyk JP, Stelmack TR, et al. Psychometric Properties of the Veterans Affairs Low-Vision Visual Functioning Questionnaire. IOVS 2004;45:3919-28.

4. Massof RW, Rickman DL, Lalle PA. Low-Vision Enhancement System. J Hopkins APL Tech D 1994;15:120-5.

5. Moshtael H, Aslam T, Underwood I, Dhillon B. High Tech Aids Low Vision: A Review of Image Processing for the Visually Impaired. Transl Vis Sci Technol 2015;4:6.

6. Culham LE, Chabra A, Rubin GS. Clinical Performance of Electronic, Head-Mounted, Low-Vision Devices. Ophthalmic Physiol Opt 2004;24:281-90.

7. Peli E, Peli T. Image Enhancement for the Visually Impaired. Optical Engineering 1984;23:47-51.

8. Crossland MD, Silva RS, Macedo AF. Smartphone, Tablet Computer and E-Reader Use by People with Vision Impairment. Ophthalmic Physiol Opt 2014;34:552-7.

9. Vargas-Martin F, Peli E. Augmented-View for Restricted Visual Field: Multiple Device Implementations. Optom Vis Sci 2002;79:715-23.

10. Zhao Y, Hu M, Hashash S, Azenkot S. Understanding Low Vision People's Visual Perception on Commercial Augmented Reality Glasses. CHI Conference on Human Factors in Computing Systems 2017:4170-81.

11. Hwang AD, Peli E. An Augmented-Reality Edge Enhancement Application for Google Glass. Optom Vis Sci 2014;91:1021-30.

12. Al-Atabany W, Memon MA, Downes SM, Degenaar PA. Designing and Testing Scene Enhancement Algorithms for Patients with Retina Degenerative Disorders. BioMedical Engineering OnLine 2010;9:27.

13. Peli E, Goldstein RB, Young GM, et al. Image Enhancement for the Visually Impaired. Simulations and Experimental Results. IOVS 1991;32:2337-50.

14. Peli E. Vision Multiplexing: An Engineering Approach to Vision Rehabilitation Device Development. Optom Vis Sci 2001;78:304-15.

15. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Edition. In: World Health Organization; 2016.

16. Hicks SL, Wilson I, Muhammed L, et al. A Depth-Based Head-Mounted Visual Display to Aid Navigation in Partially Sighted Individuals. PLoS One 2013;8:e67695.

17. Bordier C, Petra J, Dauxerre C, et al. Influence of Background on Image Recognition in Normal Vision and Age-Related Macular Degeneration. Ophthalmic Physiol Opt 2011;31:203-15.

18. Everingham M, Thomas B, Troscianko T. Head-Mounted Mobility Aid for Low Vision Using Scene Classification Techniques. Int J of Virt Reality 1998;3:1-10.

19. Everingham MR, Thomas BT, Troscianko T. Wearable Mobility Aid for Low Vision Using Scene Classification in a Markov Random Field Model Framework. International Journal of Human-Computer Interaction 2003;15:231-44.

20. Peyrin C, Ramanoel S, Roux-Sibilon A, et al. Scene Perception in Age-Related Macular Degeneration: Effect of Spatial Frequencies and Contrast in Residual Vision. Vision Res 2017;130:36-47.

21. van Rheede JJ, Wilson IR, Qian RI, et al. Improving Mobility Performance in Low Vision with a Distance-Based Representation of the Visual Scene. IOVS 2015;56:4802-9.

22. Wilson IR, Van Rheede JJ, Campbell AM, et al. A Mobile Image Enhancement System for Sight Impaired Individuals. IOVS 2016;57.

23. Microsoft. Hololens Hardware Details. https://developer.microsoft.com/en-us/windows/mixed-reality/hololens_hardware_details. Accessed: 07.11.2017.

24. Odell NV, Leske DA, Hatt SR, et al. The Effect of Bangerter Filters on Optotype Acuity, Vernier Acuity, and Contrast Sensitivity. J AAPOS 2008;12:555-9.

25. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977;33:159-74.

26. R Version 3.3.2: Vienna, Austria: R Foundation for Statistical Computing; 2016. Available at: https://www.R-project.org.

27. Derrick B, Russ B, Toher D, White P. Test Statistics for the Comparison of Means for Two Samples That Include Both Paired and Independent Observations. J Mod Appl Stat Meth 2017;16:137-57.

28. Bochsler TM, Legge GE, Kallie CS, Gage R. Seeing Steps and Ramps with Simulated Low Acuity: Impact of Texture and Locomotion. Optom Vis Sci 2012;89:E1299-307.

29. Kallie CS, Legge GE, Yu D. Identification and Detection of Simple 3d Objects with Severely Blurred Vision. Invest Ophthalmol Vis Sci 2012;53:7997-8005.

30. Wang SH, Pan HR, Zhang CY, Tian YL. Rgb-D Image-Based Detection of Stairs, Pedestrian Crosswalks and Traffic Signs. Journal of Visual Communication and Image Representation 2014;25:263-72.

31. Mamassian P. Visual Confidence. Annu Rev Vis Sci 2016;2:459-81.

32. DeCarlo D, Finkelstein A, Rusinkiewicz S, Santella A. Suggestive Contours for Conveying Shape. ACM Transactions on Graphics (Proc. SIGGRAPH) 2003;22:848-55.

33. Rusinkiewicz S, Burns M, DeCarlo D. Exaggerated Shading for Depicting Shape and Detail. In: ACM Transactions on Graphics; Boston, MA: Boston, MA. ACM; 2006:1199-205.

Figure legends

Figure 1. HoloLens hardware. **(A)** The HoloLens has binocular see-through displays (red arrows), a sensor bar (blue dashed box), and an onboard computer. **(B)** Users wear HoloLens by tightening an adjustable band around the head and positioning the screen in front of their eyes.

Figure 2. Augmented reality application. **(A,B)** The HoloLens creates a 3D map of the physical environment (A) and can overlay an augmented stereoscopic view (B). The example overlay shows nearby surfaces (less than 2.0 m) as warm colors, and farther surfaces as cool colors (2.0 m and farther). This is the red-to-blue AR used in Experiment 2. **(C)** Several other example views of the same scene demonstrate how the color and intensity can be customized for individual users. These are a subset of the options presented to participants in Experiment 1. All examples were captured from the HoloLens using the scene camera positioned between the user's eyes.

Figure 3. Example tasks. (**A**) Images of the five objects. To decrease the probability of getting the correct answer based purely on the approximate size of the object, several objects were selected to have a similar shape and size. (**B**) Example of one of the three unique room layouts used in the mobility task for Experiment 2. Layouts were comprised of a set of tables and chairs in different locations, with different objects placed on the tables as well.

Figure 4. Experiment 1 results. Results are shown for person localization **(A)**, pose recognition **(B)**, object recognition **(C)** and mobility **(D)** tasks. Bar heights in A-C indicate percent correct (left column) and average confidence (right column) of each participant across baseline and AR trial blocks. Bar heights in D indicate mean distance each participant stopped in front of the obstacle in the two trial blocks. On the last AR trial, Participant 4 changed direction and walked towards a wall. He detected the wall visually using the AR color before hitting it with his cane, so this distance was recorded and used for analysis. Error bars in confidence ratings and mobility task indicate standard error. AR = Augmented Reality; p1-4 = Participants 1-4.

Figure 5. Experiment 2 results for visual identification tasks in terms of percent correct (**A**, **B**, **C**) and confidence ratings (**D**, **E**, **F**). Bar heights indicate the mean across participants within each group (control, color, opacity), and error bars indicate standard error. Results for each group are summarized with two

bars that represent data from the first (baseline) and second (AR) block of trials. AR = Augmented Reality. *** P < .001; * P < .05.

Figure 6. Relationship between performance and confidence. Mean confidence ratings are shown separately for correct/incorrect trials in pose (**A**), object (**B**), and gesture (**C**) recognition. Data are plotted as in Figure 4, except the two test conditions (color and opacity) are combined. AR = Augmented Reality. Partially overlapping t-tests were used to compare means between binned data. T-statistics (degrees of freedom) for significant differences: Pose recognition, correct trials, test condition: T(121.7)=3.08; Object recognition, correct trials, test condition: T(112.15)=4.81; Object recognition, incorrect trials, test condition: T(144.26)=5.49; Gesture recognition, correct trials, test condition (unpaired t-test): T(17) = 3.69; Gesture recognition, incorrect trials, test condition: T(144.26)=-5.49. *** P < .001; ** P < .01, * P < .05.

Figure 7. Experiment 2 mobility task results. Differences in subjective responses compared to baseline in the control group and test groups are shown for risk of collision **(A)** and usefulness of vision **(B)**. Positive values indicate ratings higher than baseline, and the maximum absolute difference is 6. Error bars indicate standard error. AR = Augmented Reality. *** P < .001; ** P < .01.