

Attention Guiding Techniques using Peripheral Vision and Eye Tracking for Feedback in Augmented-Reality-based Assistance Systems

Patrick Renner*

Thies Pfeiffer†

Cluster of Excellence Cognitive Interaction Technology
Bielefeld University
Inspiration 1, 33619 Bielefeld, Germany

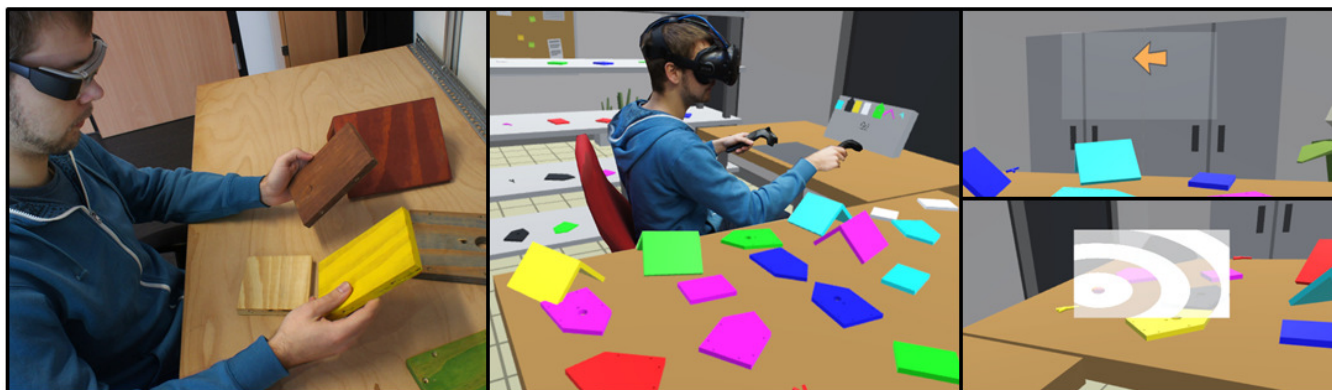


Figure 1: We are targeting the problem of attention guiding in assembly and picking tasks (left). To be able to systematically and device independently evaluate own designs (bottom right) and established designs (e.g. arrow, top right), we simulate the application context and the AR device in virtual reality (middle).

ABSTRACT

A limiting factor of current smart glasses-based augmented reality (AR) systems is their small field of view. AR assistance systems designed for tasks such as order picking or manual assembly are supposed to guide the visual attention of the user towards the item that is relevant next. This is a challenging task, as the user may initially be in an arbitrary position and orientation relative to the target. As a result of the small field of view, in most cases the target will initially not be covered by the AR display, even if it is visible to the user. This raises the question of how to design attention guiding for such "off-screen gaze" conditions.

The central idea put forward in this paper is to display cues for attention guidance in a way that they can still be followed using peripheral vision. While the eyes' focus point is beyond the AR display, certain visual cues presented on the display are still detectable by the human. In addition to that, guidance methods that are adaptive to the eye movements of the user are introduced and evaluated.

In the frame of a research project on smart glasses-based assistance systems for a manual assembly station, several attention guiding techniques with and without eye tracking have been designed, implemented and tested. As evaluation method simulated AR in a virtual reality HMD setup was used, which supports a repeatable and highly-controlled experimental design.

Index Terms: H.5.2 [Information Interfaces and Presentation (e.g. HCI)]: User Interfaces—Miscellaneous

*e-mail: preenner@techfak.uni-bielefeld.de

†e-mail: tpfeiffer@techfak.uni-bielefeld.de

1 INTRODUCTION

Augmented reality (AR) glasses have promised to be the tool for the smart worker of the future (cyberphysical systems, Industrie 4.0) for a long time. Especially for assistive systems in the areas of maintenance, production, and health care, AR technology is already being actively used: If the user wears AR glasses instead of holding a tablet or a notebook for getting necessary information, his hands are free for working on the task. It has been shown that the use of AR reduces head movements and decreases error rate in manual tasks [29]. Video-based AR systems however often come with a limited total field of view, which has an impact on task performance [33]. Optical see-through systems have typically a smaller field of view for AR content than video-based systems, yet they do not obstruct the natural field of view of the user so much.

This research is part of the larger research project "Adaptive and Mobile Action Assistance in Daily Living Activities" (ADAMAAS) with a focus on mobile assistance systems for manual assembly tasks using optical see-through smart glasses with support for eye tracking. The project's central research question addresses the personalization of such an assistance system. The system monitors the progress of the user in the task and then prompts the next component to be picked up. In this context, this paper presents work on identifying and designing attention guiding mechanisms which most efficiently direct the users' attention towards the next target under the constraints of the restricted field of view of an optical see-through AR system with a partial overlap of the user's field of view.

Many assistance functions in the aforementioned application areas require guiding the user's attention to a specific object or area of interest, which can be anywhere around him. Hence the target might not necessarily be in the field of view of the user (within view vs. beyond view); it might also not be within the field of view covered by the display area of an optical see-through AR device

(within AR view \subset within view). While there are yet a number of assistive systems, there has been only little research on how to solve this issue most efficiently.

The problem that optical see-through AR devices only cover a small field of view (most provide $< 30^\circ$ horizontally) when compared to the field of view of a human ($> 180^\circ$ horizontally) still holds even for devices that have been presented in 2016 or announced for 2017. The Microsoft HoloLens supports about 35° diagonally [31], the ODG R-7 30° [11]. The Meta 2 supports up to 90° diagonally [19], but on the costs of a lower resolution of 20 pixels per degree, as compared to the about 45 pixels per degree of the Microsoft HoloLens. The Magic Leap system is supposed to provide a large field of view and a high resolution, but no precise information has been made available at the time writing.

The most obvious criterion for a successful attention guiding technique is the speed with which the cued target is attended to. However, there are other important factors that have to be considered. First of all, it is preferable that the attention of the user is primarily directed towards the environment and not towards the AR display. Many situations in natural environments are dynamic and potentially hostile, in particular, when the user is already handicapped. Attending to an AR display while walking, for example, increases the risk of stumbling and falling. Second, a method that requires less head movements is preferable. Many techniques require a scanning-like movement with the head to cover the search space in the environment. The smaller the field of view of the AR display, the more scanning movements have to be made. On the one hand, this will quickly get tiresome, and on the other hand, especially elderly people may be less flexible in the range they can rotate their heads and would then either have to rotate their body or simply have AR-blind spots they cannot reach, e.g. the upper boards in the kitchen.

The outline of the paper is as follows: first, a review of existing work on visual guidance for AR and on the VR-based method for conducting the experiment is given. Then the proposed new approach together with different alternatives from literature are presented in detail. After that, the paper reports on a study with 20 participants in which the different approaches have been evaluated. The paper is wrapped up by a subsequent summary.

2 RELATED WORK

This work addresses one main research question: How can a technical system guide the attention of a user towards a specific target object in the environment? In the domain of AR glasses, this research problem is one of attention management and augmentation in AR interfaces [4]. Attention guidance is relevant in many application areas, such as virtual teleconferences, visual search (e.g. order picking [25, 21]), spatial navigation, or procedural queuing. The work at hand focuses on the latter: the guidance of the attention of a worker towards relevant parts during an assembly procedure. The case example is that of constructing a bird house.

2.1 Applicability of AR Glasses for the Application Domain

Typically, AR assistance applications are applied in areas like training, maintenance and construction. E.g., in a disassembly task, workers can be provided with augmented instructions to solve their task more efficiently [1]. In the area of maintenance, relevant parts can be highlighted [3] and propose different sets of instructions [27].

AR glasses may be used to present task relevant information within view of a worker and by this means reduce head movements towards classical instruction material. Ideally, this will reduce the time on task. However, perceptual issues, such as occlusions of the target objects by virtual information overlays or a bad readability of the information presented on the display over a cluttered

background of objects has reportedly a negative impact on performance [29]. The authors also provide results suggesting that the presentation of information on an AR display reduces errors and the cognitive load of participants. They also report that the presentation of information at the center of vision was perceived to be problematic by the participants, however, the technology used in their work did not allow for off-center information display.

Khuong et al. [13] apply two AR visualization methods for supporting assembly tasks using real-time detection of the assembly status. They found that displaying guidance information as an overlay to the physical target leads to longer completion times than displaying information in a side-by-side manner adjacent to it. Another way to provide the user with real-time information is using previously recorded actions and projecting that material at the corresponding time into the scene [18].

Besides smart glasses, projection-based approaches for guiding attention towards the relevant objects in a working environment are well established and successful [23, 10]. They come with the advantage that no glasses have to be worn or calibrated. The work presented in this paper, however, is part of a larger project on building personalized assistance systems for everyday life and working activities. Smart glasses would allow for a better personalization and could thereby help to address better in particular those people, who would not benefit or even be impeded by general guidance systems [16].

2.2 Design Space for Visual Techniques for Attention Guidance

As it turns out, AR glasses promise many opportunities for an optimization of working processes, yet they also provide a huge and still largely unexplored design space. Some of the key aspects of this design space will be brought into focus in the following:

In-Situ vs. In-View information display: AR glasses may present information either two-dimensional on the canvas of their display (in-view) or registered to the 3D position of the target object (in-situ). Combinations are also possible, e.g., by projecting the 3D position of the object on the 2D canvas of the display and presenting the information there.

Field of view: The display area of typical AR glasses covers a field of view of 20° to 60° [5], only recent prototypes, such as Meta 2, provide field of views of 90° . In contrast, the full field of view of a human is about 180° . Hence, a large part of the natural field of view is not covered by the AR display. Thus in the past, visualization techniques have been optimized for small fields of view. Kishishita et al. [14] show that larger fields of view increase performance in visual search tasks. Large-field displays provide better opportunities to bring digital labels closer to the location of targets. They also provide more space to arrange labels, leading to less information clutter. However, such large-field AR glasses are not yet commonly available on a broad scale.

A review of further perceptual issues is provided in Drascic and Milgram [7] and more recently in Kruijff et al. [17].

2.3 Approaches for Visual Attention Guidance

A straight-forward way of guiding the attention towards an item in the environment is highlighting [9]. In its basic form, the outlines of the object are emphasized using a 3D registered overlay. This technique is naturally constrained to the size of the AR display and thus only applicable in a small field of view, thus the user has to move his head in order to find the highlighted target in the AR display.

In the same paper [9], Feiner et al. used a label within the AR-view and an associated line in the style of a dotted rubber band to connect to targets outside the AR-view.

A different approach to guide attention towards a target is by prompting the direction towards the target using an arrow. This has been described for pedestrian navigation [8] and for guiding tourists

to points-of-interest [24]. It has also been shown to be efficient in the case of directing drivers of a car [30] using a 3D arrow. A fixed-screen 2D arrow has also been used to help mechanics to re-orient towards targets behind them [12]. As a variant of showing an arrow at a certain position of the screen, a well-known technique from gaming interfaces is positioning it at the border of the screen in direction of the target object.

As an alternative to arrows, Baudisch and Rosenholtz [2] propose the Halo technique for guiding attention to targets located outside a 2D screen: They surround off-screen targets with rings of a radius that reaches into the border region of the screen. This way, not only the location is indicated, but also the distance can be approximated based on the size of the visible arc. A limitation of the Halo technique as well as 2D arrows is that they can only indirectly provide information about targets behind the user. Thus, are not really well-suited for 360°-applications.

The omnidirectional attention funnel (OAF) is an animated visual guiding system, in which a flexible tunnel of frames is drawn from the current head position and orientation to the intended position and orientation when facing the target [4]. In a comparison study against audio cuing by naming and selection-box highlighting, the OAF could improve search assistance performance in terms of shorter search times, lower errors, and a lower cognitive load. For their study, the authors used a similar set-up as the one used in the present work, however, the task took place in the real world and not in a virtual simulation.

In real-life conditions, a combination of several techniques might be adequate, e.g. as described in [12]. They use a 2D arrow for extreme angles in which the user is required to turn around, then switch to 3D arrows when the orientation change is less than 90° and once the target is covered by the AR-view slowly migrate to a highlighting, which in turn is only active for five seconds.

While especially the latter combination of techniques appears to be very plausible and was well accepted by the participants in their study, to our knowledge there do not exist extensive reports comparing and covering the individual visualization strategies, even less covering combinations of designs. A better knowledge of the advantages and disadvantages of different visualizations, however, is necessary to support an adaptation of the system to user preferences and needs, e.g. if individuals have problems with stereoscopic perception, are color-blind, or far-sighted, just to name a few. In addition to that, certain styles of visualization also come with increased demands regarding display hardware or tracking performance, which are also constraints to be considered when designing an assistance system. Beyond that, it has been shown that too much information displayed on an AR device might degrade the awareness of users for unexpected events [6], thus a careful and informed consideration of the choice of visualization seems mandatory in critical situations.

In this paper, we will thus focus on individual guiding strategies and their performance in relation to different distances of the target from the current viewing position.

2.4 Eye Tracking for AR Assistance

Due to recent developments, the combination of AR glasses and eye tracking is commercially available (Sensomotoric Instruments' upgrade for Epson Moverio BT-200 [26]). For the same device, pupil labs also offer an eye tracking kit [20]. By the integration of eye tracking into AR glasses, several problems can be addressed: first of all, the calibration of the AR glasses can be optimized, as the devices have to detect the position of the eyes and their orientation, which is relevant for creating the correct visual perspective. Second, eye gaze might be used as an input modality for interaction with the device. And, finally, the AR glasses may monitor the direction of gaze and modulate their feedback accordingly. The latter approach is followed in this paper by some of the attention guiding techniques presented in the next section.

2.5 VR Simulation of AR Systems

When testing augmented reality systems, there are many technical aspects that make reliable studies difficult. First, in particular with optical see-through systems, there is the problem of maintaining constant environmental conditions, e.g. lighting, for each participant. Second, if repetitions of manual interaction tasks with randomized placements are required, it takes time to re-arrange parts between tasks without the participant watching. This increases overall study-time and raises problems if the re-arrangements are made inconsistently with the design due to human errors. Third, when using typical vision-based AR systems there is always the possibility of tracking errors, latencies, or other technical problems resulting in perceptual errors. Some of these errors might not be noticed by the experimenter. All of these errors, however, will influence the outcome of the study. Fourth, the general performance of the available hardware will also influence the outcome, e.g. a high tracking or display latency will severely affect the evaluation of the displayed information.

To tackle these issues, we developed a system which can abstract from hardware and environmental factors. It is capable of simulating AR devices in virtual reality (VR). The 'real world' content is realized like a normal VR simulation, but additionally AR hardware is also simulated. This makes it possible to evaluate AR techniques for arbitrary devices and completely independent of changes in the environment. Different field of views (FOVs), tracking latencies etc. can also be simulated and thus be systematically tested.

Earlier work following similar ideas has focused on specific systemic aspects and not on experiments [32]. The simulation of AR displays has been brought into focus again recently in studies on the effects of a wide field of view [15, 28].

3 VISUAL ATTENTION GUIDANCE TECHNIQUES

Usually, techniques for guiding the user's attention towards the position of the relevant item can be separated into two classes. One is displaying information "in-view" of the AR display, i.e. it moves with the user's head position and orientation. The other is "in-situ". In this case, information is directly registered at the relevant item. Here, information can be invisible if outside the FOV of the AR device. Additionally, information can be static (at a fixed position) on the AR display, or it can be adjusted relative to the position and orientation of the target item.

In this paper we propose attention guiding techniques which do not require the user to focus on them to perceive their information. The idea is that while the user's focus of attention remains on objects in the environment, the informational pattern of the visualization can still be recognized using peripheral vision. This way, the user will be able to maintain attention on the task while still being able to use relevant assistance functions. One key advantage would be that no refocusing on the AR display is required.

In prior research, we compared basic visualizations for attention guidance [22]. Building on the insights collected there, we chose combinations of visualizations as described in the following. In all cases, an in-situ highlight in form of a proxy box is shown at the target item. This proxy object is registered with the target in 3D space and is dynamically updated whenever the user moves his head. This requires tracking of the head. Constrained by the AR-view, the highlight can only be seen when the AR-view at least partially overlaps with the target.

3.1 Established Guidance Techniques and Baseline

In the following a set of three guidance techniques is presented, each of which is a realization of a well-known approach, which was implemented to compare the newly proposed techniques to.

Stationary Display as Baseline (IMAGES)

This very simple attention guiding system does not make use of the AR glasses at all. Instead, a standard computer screen located beyond the construction area will show a sequence of images of the parts to be picked up. In the target task of creating a bird house, all seven parts of the task were presented simultaneously in the order of the tasks from left to right (Figure 2 a)). The display was on until the completion of the task. The user could either memorize the parts or orient herself towards the display to see which part is next to be picked up. No AR visualizations were shown at all, however, the simulated AR display was in place, so that the slight visual degradation was held constant over all conditions. This guidance technique was used as a baseline in the study.

Attention Funnel (FUNNEL)

We evaluated the well-known attention funnel as proposed by Biocca et al. [4] (Figure 2 b)). They found that in search tasks, this visualization technique significantly reduces mental workload and visual search time in comparison to highlighting and audio cues. As the attention funnel (to the best of our knowledge) was only tested against audio cues and 3D highlights. We include it in our evaluation in order to compare this well-known technique to a broader range of visual guidance methods. This way, we also investigate the benefit of the attention funnel for small FOVs.

In-View Arrow-Based Guidance (ARROW)

An arrow is shown at a fixed position of the AR display. We chose this variant over an arrow dynamically moving at the border of the screen as we did not expect differences in performance. The orientation of the arrow is updated in a way that it always points towards the target object (Figure 2 c)). Thus, the user has to follow the direction of the arrow until the target item overlaps with the AR display. Then the additional 3D highlight is shown. Here, the additional 3D highlight is important to disambiguate between objects at different depths.

3.2 Newly proposed Guidance Technique: Spherical Wave-Based Guidance (SWAVE)

This method is the first of the newly designed methods presented in this paper. The basic idea is similar to the Halo technique [2], but expands it to 3D and uses the whole FOV of the AR display. The core idea is that of waves propagating towards the target, like the concentric circles that appear when throwing a stone in the lake, only with the inverse direction. To realize this visualization, a sphere is centered at the head position of the user. The radius is updated with respect to the distance from the user's head to the target position in a way that the target is always on the surface of the sphere. Using a graphics shader, spherical waves are rendered on the surface of the sphere, propagating towards the target.

This visualization combines in-situ and in-view aspects: The spherical waves are converging in 3D at the target item, by this also emphasizing the distance in depth for target. At the same time the waves are always in view of the AR display, providing the user with directional information about the location of the target both by their curvature (the target is located towards the concave part of the arc) and their direction of movement. A dedicated highlighting of the target is in principle also no longer needed (as long as the waves are always shown). Figure 2 e) shows an example of this guidance technique in the AR display, Figure 2 f) shows the created sphere from the outside.

The wave guides are easy to implement using a shader. They can also be implemented as a transparent screen space effect on the AR display without the need for an explicit 3D representation (then running completely in-view). As the wave guides have similar properties as the combination of an arrow and 3D highlighting, a similar performance, if not better, is expected. Advantages are that

the speed and size of the wave arcs visible in the display encode the distance, information which is not present in the arrow, and that the wave guides provide a single homogeneous concept for all distances to the target. The waves can graphically be designed to have a high contrast, which, together with their movement, should make them easily detectable in the periphery. We chose the angular speed of the waves to approximately $5^\circ/s$ for this study.

3.3 Adaptive Guidance using Eye Gaze Information

A second idea followed in this paper is that of adapting the guidance depending on the current direction of eye gaze, or more precisely, the current distance of the eye gaze towards the target. This idea has been guiding the following three designs.

Peripheral Flickering of the AR Display combined with Stationary Display (IMAGES_ET)

Building on the stationary display visualization, the AR display is providing information if the user is looking (with the eyes, not with the center of the AR display) at the correct item or is looking close to it. This is realized by making the AR screen flicker between white and transparent if the eye gaze of the user has a small angle to the target. If the target is fixated, the screen stays constantly white. If the user looks far away from it, the screen stays transparent. If the AR screen overlaps with the target, the full-screen flickering is reduced to a frame, preventing the target from being occluded.

When flickering, the screen is white for 40 ms, thus a very short but still actively perceivable duration. We interpolated the frequency of the flickering between 1 Hz and 5 Hz to make it notable but not obtrusive.

In-View Arrow combined with Peripheral Flickering (ARROW_ET)

This combination is designed to improve upon the standard arrow-based guiding. By monitoring the eye gaze, visual feedback similar to the previous technique by peripheral flickering is given once the eye is approaching the target. By this means the user should be ensured with their decision and bringing the AR display to overlap with the target should no longer be necessary for target validation. An example of the AR display when flickering white is shown in Figure 2 d).

Spherical Wave-Based Guidance with Dynamic Speed of Movement (DynSWAVE)

The SWAVE guidance can also be extended using eye gaze. Here, the angle between the current fixation and the target item is used as a parameter which controls the speed of the spherical waves. Thus, if the user directly looks at the target object, the waves do not move anymore, while they move faster when the angle between the fixation and the target increases (up to 6-times the base speed), thus emphasizing the distance towards the target.

3.4 Hypotheses

Based on the results of previous work, we supposed the stationary display visualization to take more time for finding the correct item, as visual search for the item has to be conducted. The 3D highlight however can help the user to validate if the correct item was found.

The 2D arrow-based guidance alone is supposed to lack necessary depth information, but combined with the highlight, it should provide precise information on the 3D position of the target. The newly proposed SWAVE approach, which inherently combines in-view and in-situ information, should perform at least as well as the arrow-based approach. On the other hand, it is doubted that the attention funnel is well suited for small FOVs, as ideally, there should be some space to make this approach look like an actual funnel.

The additional peripheral cues are supposed to help users finding the target without necessarily having to overlap the AR display

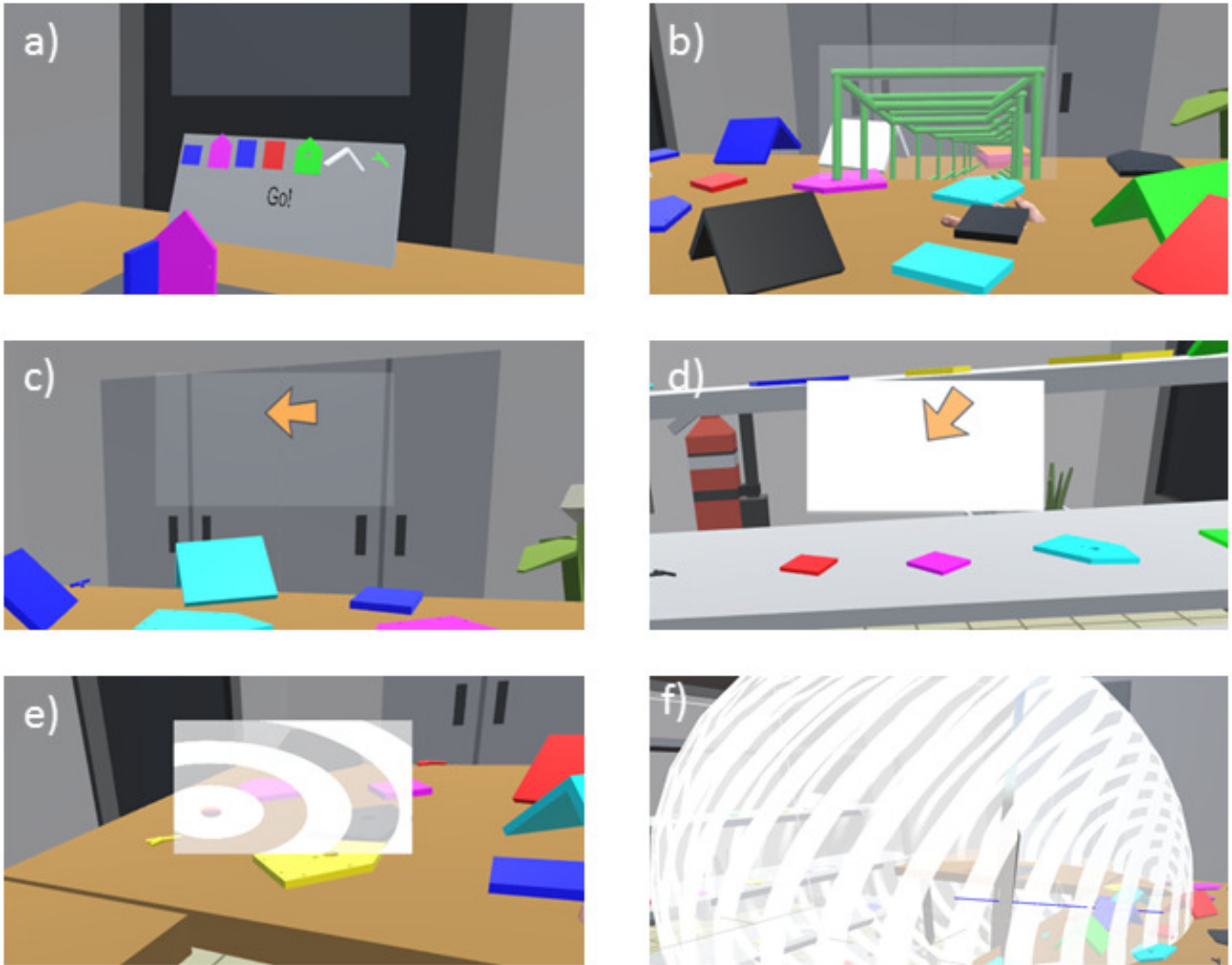


Figure 2: The conditions of the study: a) Images shown on a screen, b) Attention Funnel (3D highlight is also visible), c) 2D arrow, d) Example of the white display when flickering (here in the arrow-based guidance), e) Spherical Wave-based guidance (SWAVE), f) Realization of the wave-based guidance: Look from the outside at the sphere with the wave shader.

with potential targets. We expect the peripheral flickering to speed up time-on-task when combined with the baseline images, giving direct feedback on the target object being close. However, we do not expect this effect so much for the combination with arrow-based guidance. Instead, this combination is supposed to reduce fixations on the AR display. In the SWAVE guidance, dynamically changing the speed of the wave movement should provide the user with more precise information on the distance to the object. This way, it should decrease search time in particular when the target is further away.

4 METHODOLOGY

We conducted a within-subject experiment in a construction scenario. The seven types of assistance, as described in Section 3, were evaluated as independent variable. The conditions thus were the images on display baseline ("IMAGES"), the attention-funnel ("FUNNEL"), the images combined with flickering ("IMAGES_ET"), the in-view arrow ("ARROW"), the arrow combined with flickering ("ARROW_ET"), the spherical waves ("SWAVE") and the waves with dynamic speed ("DynSWAVE"). The distance between two

target-objects was co-varied, while the overall distance in each repetition was held approximately constant. Under these constraints, all target positions were chosen randomly.

Participants

We had 20 participants (13 male and 7 female) taking part in our experiment. Their age ranged from 19 to 27, with a mean of 22.2 years (sd: 2.2). They were all students from our university from various fields of study. Only four of them had reasonable experience with virtual reality, while 13 were experienced with computer games.

4.1 Materials

Presentation and Interaction

An HTC Vive HMD with an integrated 120Hz binocular Pupil Labs eye tracker was used for simulating VR and AR content. With a latency of 5.7 ms, the eye tracker is suitable for real-time interaction. Its gaze accuracy is 0.6° , the precision is 0.08° . Due to the HMD room-scale tracking capabilities of the HMD, study participants were able to move freely in their workspace. The two con-

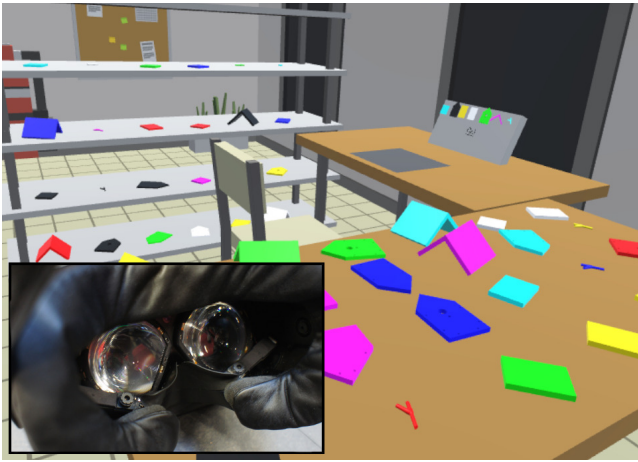


Figure 3: The scenario: Parts of the birdhouse are distributed over the left shelf and the table on the right. The inner image shows the Pupil Labs Eye Tracker which was integrated into the HTC Vive HMD.

trollers of the HTC Vive were used to visualize simulated hands which made a grasping movement when pressing the trigger knob. A number of participants tried to place their virtual hands on the VR table in front of them, which can be seen as an indication of a high immersion of the simulation.

We assumed that for users who are unexperienced in VR as well as AR, taking part in a simulated AR study needed some time to accommodate. Therefore, a training phase in the beginning allowed participants to get used to the simulated environment. Then, before the first actual trial with a new kind of visualization, these were demonstrated to the participants and explained. After that, participants were instructed to complete the task as fast as possible using the provided support.

The eye tracker was calibrated before the beginning of each condition to make sure that eye gaze information was provided with constant quality.

Scenario

A small workshop environment was simulated, where the user sat at a desk for building a birdhouse (see Figure 3). The parts to pick were located in a shelf to the left and on a desk to the right of the user. They were randomly re-arranged for every trial using a set of 48 fixed positions, under the constraint that the full path length (i.e. the sum of all angles between the parts to pick) ranged between 500° and 600° . Also, three or four of the seven objects to pick had to be located in the shelf. In addition to that, the orientations of the parts were also randomly chosen. Such a setup would not have been easy to realize in a real-world AR study.

The participants wore simulated AR glasses. Here, the field-of-view of an Epson Moverio BT-200 was simulated in relation to the FOV of the HMD. This was done in order to simulate a natural ratio between human FOV and the FOV of the AR glasses. Specifically, as the FOV of the HTC Vive HMD was only approximately 95° when using the integrated eye tracker. As this is close to half of the natural human FOV, we accordingly decreased the size of the AR FOV. Additionally, the AR display was slightly moved to the top to match the ratio between vertical human FOV and the AR FOV at least for the vertical direction. Even when there was no AR content shown, they could notice the FOV of the glasses as a rectangle which was slightly brighter than the environment. This effect also occurs in reality when wearing the Epson Moverio. This made participants always aware of wearing the glasses.

The interaction with the parts, which was not relevant for this

study, was minimized: Grasping the objects (with simulated hands) directly moved them to the construction area. This way, users could completely focus on finding the correct objects. Also, using the simulation it was possible to automatically calculate more accurate statistics about angles to and distances between the parts.

Task

For the user study we chose a picking task for a construction scenario: The birdhouse consists of seven parts, which had to be found out of 48 parts of seven colors and six shapes. The correct order of this picking task was mandatory. Half of the parts were placed in four rows of five on a table on the right of the user. The other half in four levels of five parts in a shelf on the left. Thus, half of the parts were located in the vertical plane (in the shelf) and half of them in the horizontal plane (on the desk).

4.1.1 Procedure

Initially, the user sat in front of a construction area where the birdhouse was built. Behind that area, a conventional computer display was simulated for giving instructions in the baseline condition. In order to set the focus on visual search, only picking the correct target was required (mounting the birdhouse was done automatically by the simulation).

For each AR guidance technique, five bird houses had to be built. As each only consisted of seven parts, one house was usually built in less than 30 seconds. Altogether, 35 houses had to be built. The first condition always was the baseline (showing images of the parts on a virtual screen in front of the participants) to make sure that all participants had the same knowledge about the task in the beginning. The order of the other techniques was fully randomized, with five repetitions of each technique en bloc.

Finally, after the last run, participants were asked to rank the techniques according to different criteria: Their impression about speed, accuracy, learnability, and usefulness.

During the trials, the time to pick a part of the birdhouse was measured as well as the angle and distance to the prior part. We also measured the head movement angle that was traveled during finding a part. Additionally, we measured the ratio between eye gaze on the AR display and eye gaze into the scene.

4.2 Results

Objective Measurements

The length of a full path in angle ranged between 500° and 600° by design. We therefore normalized the time needed to build a birdhouse with the path length to make the individual data comparable. Figure 4 shows the normalized times in seconds/ 100° for building a birdhouse for the different conditions. The results were statistically tested using ANOVA and Tukey's honest significance test.

The fastest guiding technique was the arrow guidance with a median of 4.8 s per 100° (sd: 1.4). Adding flickering made it a bit slower (median: 5.1, sd: 1.9). Similarly fast was the SWAVE guidance with a median of 5.2 s/ 100° (sd: 1.2). Here, using gaze information lead to slightly faster results (median: 4.9, sd: 1.9).

The baseline condition with images on the display, as well as the version with added gaze-based flickering, were significantly slower than these techniques ($p < 0.01$). The baseline needed 6 s/ 100° in median (sd: 1.6), adding flickering turned out slightly faster (median: 5.8, sd: 1.1). The attention funnel technique was a bit faster with a median of 5.7 s/ 100° (sd: 3.1), which was significantly faster than the baseline without flickering ($p < 0.01$).

After analyzing the overall times needed to accomplish the birdhouse task with the different guiding techniques, it is also interesting to compare the techniques in specific situations: If the next part to pick is visible in the AR display, visible for the participant but not in the AR display, or not visible at all. Figure 5 shows the times needed to pick one part in the different visibility conditions. When

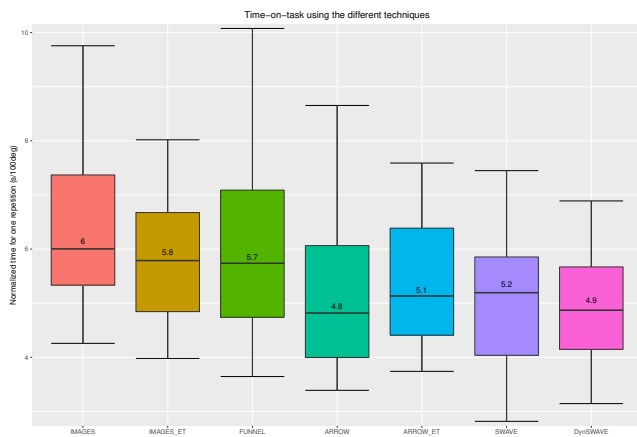


Figure 4: The times needed to build a birdhouse under different conditions. The times are normalized by the full path angle.

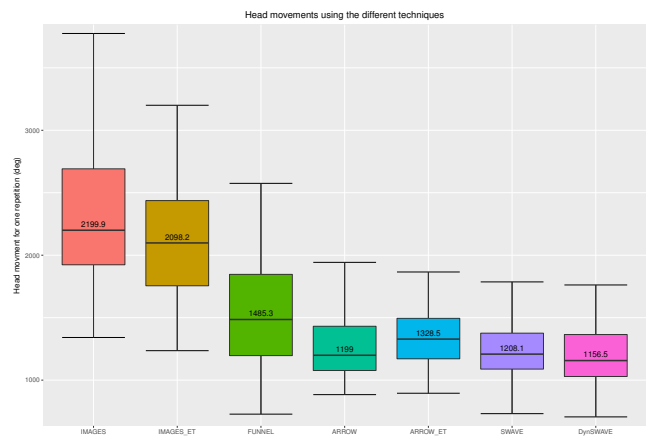


Figure 6: The head movement in degrees conducted to build one birdhouse.

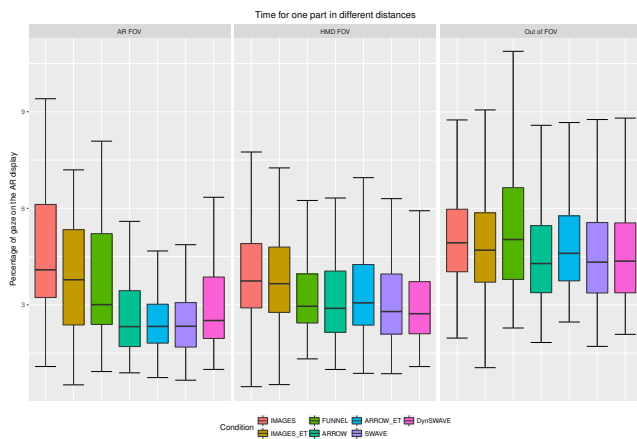


Figure 5: The times needed to pick one part, depending on its initial visibility in the HMD and the simulated AR display.

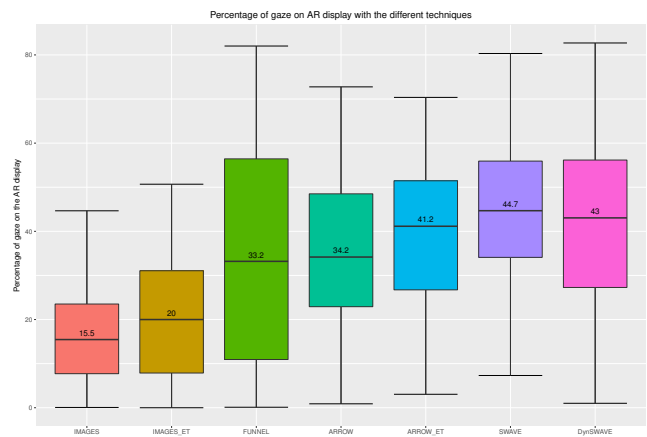


Figure 7: The percentage of looking at the simulated AR display while building a birdhouse.

the part was visible in the AR display, the arrow and SWAVE techniques were fastest with a median ranging from 2.3 s to 2.5 s per part. The baseline was slowest with a median of 4.1 s (sd: 7.1) which is significant compared to all other techniques except for the one with added flickering ($p < 0.05$). This difference vanishes for parts which were visible in the HMD, but not in the AR display. The other techniques show a trend to be faster which is not significant. The same goes for objects out of the FOV of the HMD.

Not only the time to complete the task is relevant for evaluating the techniques, but also the need to move the head to find the correct object. Figure 6 shows the overall angles needed to pick the correct parts for one birdhouse. The baseline technique needed a median of 2200° (sd: 758). With 2089° (sd: 488) adding flickering lead to a slight improvement, which is however not significant. The best technique regarding head movements was the DynSWAVE technique with a median of 1157° (sd: 347). SWAVE (without using eye gaze) as well as the arrow-based techniques lead to similar head movement angles. The funnel guidance generated more head movements (median: 1485° , sd: 734). This was significantly more than for the arrow guidance with and without flickering and the SWAVE technique ($p < 0.05$).

Finally, we investigated the percentage of gaze spent on the AR display (Figure 7). The peripheral cuing techniques did not lead to less gaze on the display. Only the image baseline as well as

the version with added flickering lead to significantly less fixations on the display as the other techniques ($p < 0.001$). Participants on average spend about 40% of the time on the display using the non-baseline techniques.

Subjective Evaluation

Figure 8 shows the subjective ratings of the guidance techniques by the study participants. They were asked to rank the techniques according to speed, accuracy, learnability and usefulness. Finally, they were also asked to choose which of the techniques they would use. The ranking of the seven techniques was transcribed as distributed points from +3 to -3 (e.g., in the case of speed, this represents "fast" to "slow").

Speed and accuracy of the baseline image-display guidance as well as the funnel guidance were rated negatively. For accuracy, the baseline with flickering and the funnel were rated less accurate than the best rated technique, which was SWAVE (without eye-gaze support). On average, it was rated with 0.75 while the least accurate technique had a rating of -0.65. However, these results were not significant.

The guidance technique that was rated as best to learn was the image baseline, which was significantly higher rated than the funnel, the arrow with flickering and DynSWAVE ($p < 0.05$). There is a general trend that the eye-tracking based techniques are harder to

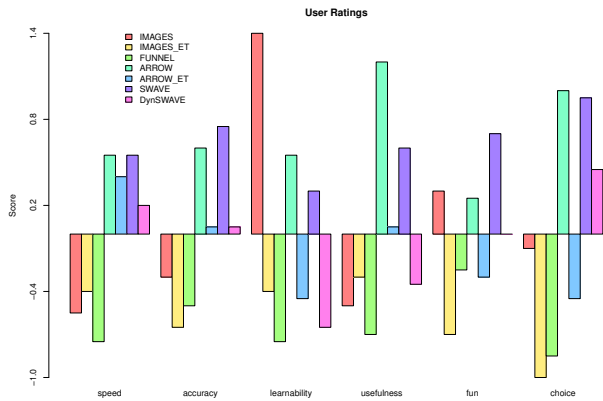


Figure 8: The subjective ratings of the techniques by the study participants.

learn than the ones without.

Regarding usefulness, participants preferred the arrow-based and wave-based guidance techniques (except for DynSWAVE). The best rated technique was the arrow-based guidance without flickering with a score of 1.2, which was significantly better than the funnel guidance ($p < 0.05$).

As the final choice, most people preferred the arrow based guidance with a score of 1.0. Also, SWAVE was preferred with a score of 0.95. Both techniques were rated significantly better than the baseline with eye-tracking and the funnel technique ($p < 0.05$). The most unpopular technique was the baseline combined with flickering. It was rated with a score of -1.0.

4.3 Discussion

Considering the hypotheses stated before, we could verify some of these, but not all. The baseline guidance technique turned out to be the slowest, as supposed. It also needed most head movements to find the correct part. As visual search had to be conducted without any assistance (as long as the part was not in the AR display), this seems reasonable. However, with a longer distance to the relevant object, this effect gets smaller. Adding eye-gaze based flickering as a hint indeed could reduce the time to find the object and reduce head movements, but only slightly. The trade-off is that participants looked at the AR display more frequently. Still, we think that this peripheral cue could be of help in tasks where the focus should not be on the AR display but a minimal assistance is required.

We combined the 2D-arrow based guidance with a 3D highlight, which made it robust also for objects that were stacked in depth. It was fast, accurate and finally preferred by the participants. The same goes for the newly proposed SWAVE guidance. The advantage of the wave-based guidance is that it can be created efficiently using a simple ripple shader. Also, it can be realized more subtle as an arrow by just having it as nearly transparent overlay. Both techniques however inherently led to many fixations on the AR display. In conflict with our hypotheses, adding eye-gaze based support by flickering or dynamically adapting speed of the waves did not reduce fixations on the display. Considering the learnability ratings, these techniques might be too complicated and users ignored the additional help. This could mean that a longer learning phase is necessary, or the design has to be altered. As hypothesized, there is at least a small trend that dynamically changing the speed of the waves leads to a shorter search time.

In comparison, the attention funnel did not turn out to be a good technique when using a small FOV display. Users had problems

following the funnel paths. Thus, this technique turned out being quite slow and was also rated as being inaccurate.

5 CONCLUSION

Based on previous work, we evaluated different AR-based guiding techniques. Each technique was combined with a 3D in-situ highlight. We proposed the new guidance technique SWAVE using spherical waves moving towards the target. Moreover, we applied eye-tracking in order to make use of the user's current gaze. We added incorporated eye gaze information in the guidance techniques by adding flickering if looking close to the target as well as by changing the speed for the spherical waves.

We evaluated the techniques using our AR simulator framework. Basically, the AR device is simulated in virtual reality which is presented in an HMD.

Altogether, the arrow-based guidance technique turned out to be fastest and best rated by the study participants. Our SWAVE guidance showed similar results. As the wave-based technique can be realized very efficiently and more subtle than the arrow, we propose it as interesting alternative to that classic approach. It could be made even more subtle by only visualizing waves at the border of the AR screen in the direction of the target, similar to the 2D Halo technique. Also, one possibility could be to change transparency of the waves instead of their speed, which would make this approach even more subtle. However, showing the arrow dynamically moving at the border of the screen would also be more subtle than our current implementation. Thus, we aim to evaluate this variant as well in the future. The attention funnel turned out to have some drawbacks when being applied in a very small FOV display as in our study.

In contrast to our hypotheses, a benefit of making use of eye gaze information could only be observed when necessary information for finding the object is given beforehand. Here, that was the case when showing images of the relevant parts on a display. Adding flickering could slightly reduce head movements and time-on-task. However, it lead to more fixations on the AR display.

5.1 Future Work

This work was part of a series of experiments in which we examine and improve AR guidance techniques. Using the insights of this study, we will try to improve the design of eye-gaze based peripheral assistance and further evaluate their benefits. Especially, we want to test different FOVs which we can easily realize using simulated AR.

Finally, we are planning to validate our results in a real-world application where a birdhouse actually has to be constructed.

ACKNOWLEDGEMENTS

This research was partly supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG) and partly by the BMBF project "ADAMAAS".

REFERENCES

- [1] H. Alvarez, I. Aguinaga, and D. Borro. Providing guidance for maintenance operations using automatic markerless Augmented Reality system. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 181–190, Oct. 2011.
- [2] P. Baudisch and R. Rosenholtz. Halo: A technique for visualizing off-screen objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03*, pages 481–488, New York, NY, USA, 2003. ACM.
- [3] B. Besbes, S. Collette, M. Tamaazousti, S. Bourgeois, and V. Gay-Bellile. An interactive Augmented Reality system: A prototype for industrial maintenance training applications. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 269–270, Nov. 2012.

- [4] F. Biocca, A. Tang, C. Owen, and F. Xiao. Attention Funnel: Omnidirectional 3d Cursor for Mobile Augmented Reality Platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1115–1122, New York, NY, USA, 2006. ACM.
- [5] O. Cakmakci and J. Rolland. Head-worn displays: a review. *Journal of Display Technology*, 2(3):199–216, Sept. 2006.
- [6] B. J. Dixon, M. J. Daly, H. Chan, A. D. Vescan, I. J. Witterick, and J. C. Irish. Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical Endoscopy*, 27(2):454–461, July 2012.
- [7] D. Drascic and P. Milgram. Perceptual issues in augmented reality. volume 2653, pages 123–134, 1996.
- [8] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. *Personal Technologies*, 1(4):208–217, 1997.
- [9] S. Feiner, B. Macintyre, and D. Seligmann. Knowledge-based Augmented Reality. *Commun. ACM*, 36(7):53–62, July 1993.
- [10] M. Funk, A. Bächler, L. Bächler, O. Korn, C. Krieger, T. Heidenreich, and A. Schmidt. Comparing Projected In-situ Feedback at the Manual Assembly Workplace with Impaired Workers. In *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, PETRA '15, pages 1:1–1:8, New York, NY, USA, 2015. ACM.
- [11] O. D. Group. ODG - System - Products. <http://www.osterhoutgroup.com/products-comparison>.
- [12] S. Henderson and S. Feiner. Exploring the Benefits of Augmented Reality Documentation for Maintenance and Repair. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1355–1368, 2011.
- [13] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. L. Viola, T. Mashita, and H. Takemura. The effectiveness of an AR-based context-aware assembly support system in object assembly. In *2014 IEEE Virtual Reality (VR)*, pages 57–62, Mar. 2014.
- [14] N. Kishishita, K. Kiyokawa, J. Orlosky, T. Mashita, H. Takemura, and E. Kruijff. Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 177–186, Sept. 2014.
- [15] N. Kishishita, J. Orlosky, T. Mashita, K. Kiyokawa, and H. Takemura. Poster: Investigation on the peripheral visual field for information display with real and virtual wide field-of-view see-through HMDs. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 143–144, Mar. 2013.
- [16] O. Korn, A. Schmidt, and T. Hörz. Augmented Manufacturing: A Study with Impaired Persons on Assistive Systems Using In-situ Projection. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments*, PETRA '13, pages 21:1–21:8, New York, NY, USA, 2013. ACM.
- [17] E. Kruijff, J. E. Swan II, and S. Feiner. Perceptual issues in augmented reality revisited. In *ISMAR*, volume 9, pages 3–12, 2010.
- [18] N. Petersen, A. Pagani, and D. Stricker. Real-time modeling and tracking manual workflows from first-person vision. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 117–124, Oct. 2013.
- [19] L. Prasuethsut. Meta 2 first impressions: AR feels closer than ever. <https://www.wareable.com/ar/meta-2-review>, May 2016.
- [20] Pupil Labs UG. Eye tracking for vr and ar. <https://pupil-labs.com/blog/2016-02/eye-tracking-for-vr-and-ar>, 2016.
- [21] R. Reif and W. A. Günthner. Pick-by-vision: augmented reality supported order picking. *The Visual Computer*, 25(5-7):461–467, Mar. 2009.
- [22] P. Renner and T. Pfeiffer. Evaluation of Attention Guiding Techniques for Augmented Reality-based Assistance in Picking and Assembly Tasks. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion*. ACM, 2017.
- [23] A. L. Sauer, A. Parks, and P. C. Heyn. Assistive technology effects on the employment outcomes for people with cognitive disabilities: A systematic review. *Disability and Rehabilitation: Assistive Technology*, 5(6):377–391, Nov. 2010.
- [24] T. Schinke, N. Henze, and S. Boll. Visualization of off-screen objects in mobile augmented reality. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 313–316. ACM, 2010.
- [25] B. Schwerdtfeger and G. Klinker. Supporting Order Picking with Augmented Reality. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '08, pages 91–94, Washington, DC, USA, 2008. IEEE Computer Society.
- [26] SensoMotoric Instruments GmbH. Eye tracking upgrade for ar glasses based on epson moverio bt-200. http://www.smivision.com/fileadmin/user_upload/downloads/product_flyer/prod_smi_eyetracking_arglasses_epson_screen.pdf, 2015.
- [27] D. Stanimirovic, N. Damasky, S. Webel, D. Koriath, A. Spillner, and D. Kurz. [Poster] A Mobile Augmented reality system to assist auto mechanics. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 305–306, Sept. 2014.
- [28] E. Steindecker, R. Stelzer, and B. Saske. Requirements for Virtualization of AR Displays within VR Environments. In R. Shumaker and S. Lackey, editors, *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, number 8525 in Lecture Notes in Computer Science, pages 105–116. Springer International Publishing, June 2014.
- [29] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative Effectiveness of Augmented Reality in Object Assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 73–80, New York, NY, USA, 2003. ACM.
- [30] M. Tonnis and G. Klinker. Effective control of a car driver's attention for visual and acoustic guidance towards the direction of imminent dangers. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 13–22. IEEE Computer Society, 2006.
- [31] A. Velvárt. HoloLens vs Meta 2 vbandi.net. <https://vbandi.net/2016/03/04/hololens-vs-meta-2/>, Mar. 2016.
- [32] A. M. Wafaa, N. D. Bonnefoy, E. Dubois, P. Torguet, and J. P. Jessel. Virtual Reality Simulation for Prototyping Augmented Reality. In *International Symposium on Ubiquitous Virtual Reality, 2008. ISUVR 2008*, pages 55–58, July 2008.
- [33] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3 edition, 2012. 04117.