AR-Glasses-Based Attention Guiding for Complex Environments

Requirements, Classification and Evaluation

Patrick Renner
Bielefeld University
Center for Cognitive Interaction Technology 'CITEC'
Bielefeld, Germany
prenner@techfak.uni-bielefeld.de

Thies Pfeiffer
University of Applied Sciences Emden/Leer
Department of Technology
Emden, Germany
thies.pfeiffer@hs-emden-leer.de

ABSTRACT

Augmented Reality (AR) based assistance has a huge potential in the context of Industry 4.0: AR links digital information to physical objects and processes in a mobile and, in the case of AR glasses, hands-free way. In most companies, order-picking is still done using paper lists. With the rapid development of AR hardware during the last years, the interest in digitizing picking processes using AR rises. AR-based guiding for picking tasks can reduce the time needed for visual search and reduce errors, such as wrongly picked items or false placements.

Choosing the best guiding technique is a non-trivial task: Different environments bring their own inherent constraints and requirements. In the literature, many kinds of guiding techniques were proposed, but the majority of techniques were only compared to non-AR picking assistance.

To reveal advantages and disadvantages of AR-based guiding techniques, the contribution of this paper is three-fold: First, an analysis of tasks and environments reveals requirements and constraints on attention guiding techniques which are condensed to a taxonomy of attention guiding techniques. Second, guiding techniques covering a range of approaches from the literature are evaluated in a large-scale picking environment with a focus on task performance and on factors as the users' feeling of autonomy and ergonomics. Finally, a 3D path-based guiding technique supporting multiple goals simultaneously in complex environments is proposed.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Empirical studies in HCI**.

ACM Reference Format:
Patrick Renner and Thies Pfeiffer. 2020. AR-Glasses-Based Attention Guiding for Complex Environments: Requirements, Classification and Evaluation. In *The 13th PErvasive Technologies Related to Assistive Environments Conference (PETRA '20), June 30-July 3, 2020, Corfu, Greece.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3389189.3389198

PETRA '20, June 30-July 3, 2020, Corfu, Greece

ACM ISBN 978-1-4503-7773-7/20/06…$15.00
https://doi.org/10.1145/3389189.3389198

1 INTRODUCTION

For applications in areas such as picking, assembly, maintenance and also in non-industrial areas, such as health-care and education, Augmented Reality (AR) can link digital information to physical objects and processes in a mobile and, in the case of AR glasses, hands-free way.

The localization and identification of objects or places in an environment is an essential task relevant for many work processes, which can be supported by an Augmented Reality (AR)-based assistance system. This will reduce the time needed for visual search and keep errors, such as wrongly picked items or false placements, low. The design of rapid attention guiding techniques is thus one area of research in augmented assistance. In tasks with up to 300 picks per hour, each second counts.

The design of attention guiding techniques is facing several challenges. Important factors are, for example, hardware constraints, such as the placement of the AR display within the field-of-view (FOV) of the user, its coverage of the FOV, its resolution and contrast, and the maximum opacity that can be achieved.

Especially the relatively small FOV that can be found in optical see-through AR glasses currently available is a hard constraint (most devices only cover less than 40°, e.g. the Microsoft HoloLens with 35° diagonally): even when relevant target objects or locations are visible in the users FOV, the AR display may not be overlapping with the target direction and thus no visual information can be displayed close to the target.

Another source of challenges is the target environment. If, for example, target locations are distributed in 360° around the user, targets can be behind the user. And if the target environment is large or crowded, target locations will often be hidden behind objects or even walls. Attention guiding techniques should thus provide users with feedback towards remote (beyond the AR display) locations.

Various approaches of attention guiding for AR glasses have been developed in the past and were evaluated in different scenarios (e.g. [7], [28],[12]). In the present paper, we will add to this line of research a comparison of well-known and new guiding techniques in a realistic picking scenario, a large-scale environment with target occlusions. Handling the important and common challenge of occlusions in large environments is, to the best of our knowledge, underrepresented in scientific research so far. For this scenario we have improved our own variant of a line-based guiding technique based on Catmull-Rom splines. We compare this with the classic highlighting [8] approach and the recently published

SWave approach [24], which showed similar performance as the standard arrow-based approach in a comparison.

For the realization, we chose an area of our university library which is structured like a warehouse, containing multiple shelves to navigate through when searching for specific books.

This paper is structured as follows: In the next section, a review of the state-of-the-art in attention guiding is given. Then, a detailed explanation of the chosen attention guiding techniques and the picking scenario is given. Finally, the paper reports on the actual experiment and is wrapped up by a discussion and a summary.

2 RELATED WORK

AR technologies are successfully applied to assist different tasks in picking, assembly and maintenance scenarios.

In maintenance tasks, e.g., relevant parts can be highlighted [3] or different sets of AR instructions can be chosen [29]. In the area of assembly, the construction process can be monitored in real-time and instructions can be given adapted to the current status [10]. This showed to improve both speed and accuracy of the task. While Funk et al. [9] found that projected in-situ contour visualizations could increase performance and reduce cognitive load of impaired workers in an assembly task, the results of Khuong et al. suggest that side-by-side visualizations perform better than in-situ visualizations [15]. Another way of giving instructions for assembly is overlaying video information with the environment. Kraut et al. [16] propose video instructions for collaborative bike assembly which outperformed figure-based manuals in their evaluation. Petersen et al. [19] use previously captured videos as in-situ overlay over the scene.

2.1 Attention guiding in workspace-like scenarios

In its most basic form, the user's attention can be guided in static scenarios where the user does not have to re-orient to see the target. All possible locations are in front of the user, e.g. on a working desk. For the assembly of different parts on a workspace, the user is standing in front of the it and only has to be guided towards the next relevant part which is located in a specific area of this workspace.

A manifest solution is using a highlight. This is sufficient to quickly find the relevant object or objects. Showing different targets at the same time is possible by simply using multiple highlights. Feiner et al. [8] realized highlights in form of an in-situ outline on the example of printer maintenance. Similar approaches showing outlines are often used to give instructions in the context of maintenance and assembly: For assisting people in everyday maintenance tasks, an AR repair guidance system was developed by Bhatia and Vijayakumar [4]. They propose their own tracking algorithm for visualizing in-situ instructions on the device to repair. On the example of a mobile phone, they highlight parts to move with an outline and show an arrow in the direction they should be moved. Alvarez et al. [1] created a framework giving disassembly information by showing animated outlines projected on the system to maintain. A very similar approach is applied by Khuong et al. [15] for assembly. The object which should added to the assembly next is indicated by an animated wireframe moving towards the assembly point. For training workers in maintenance tasks in industrial settings, Besbes et al. [3] show instructions by overlaying relevant parts in a semi-transparent color. A multi-modal solution is the Cognition-based interactive Augmented Reality Assembly Guidance System (CARAGS) [32]. It generates augmentations according to the user's current cognition state and the task progress using AR glasses and a tactile vibration device attached to the arm. AR guidance is divided in augmentation of a virtual object, on-site assembly simulation (using virtual objects) and 3D dynamic assembly paths. Arrows connect target and destination of an assembly. Tools which have to be used are highlighted by a red dot. The user can operate the system by doing hand-gestures. In an evaluation the system is compared to an AR assistance system using AR glasses for the visualizations, but mouse and keyboard for input, as well as to a classic screen-based system. They found that the proposed system is most intuitive, easy to use and satisfactory.

In-situ hightlights are only visible when the AR display at least partially overlaps with the target. If target objects can also be located at the side or back of the user, conveying the necessary orientation information is required for efficient guiding, e.g. in form of an arrow pointing towards the target location [14, 27, 31]. Henderson and Feiner [13] combined several basic attention guiding strategies in a prototype for assisting soldiers with common tasks using a tracked HMD. For extreme angles, they show 2D arrows, for smaller angles, they show 3D arrows, and when the target object enters the area of the AR display, the arrows fade out to a highlight. Finally, all visualizations are removed for occlusion-free interaction with the target. The prototype was tested against showing information on a screen or a HUD. The AR system could reduce time-on-task as well as head movements compared to the screen. Participants reported that the prototype was intuitive and satisfying for the task.

In a previous study, we compared different basic attention guiding techniques. In an environment where the targets were arranged in 360° around the user with no occlusions were given [26]. In that first study, highlights and in-view 2D arrows were evaluated. The fastest performance was achieved using an in-situ direct line towards the target. In a second study [24], we presented a new approach called SWave based on concentric circles, inspired by the HALO technique by Baudisch et al. [2]. This technique is able to convey positional and orientational cues, as well as some information on the distance to the target. In that scenario, we found comparable good performance for arrows and SWave, surpassing that of Funnel and image-based approaches.

2.2 Attention guiding picking/warehouse scenarios

In large-scale scenarios, like guiding in warehouses, the user has to navigate around shelves to find the target object. Such scenarios require attention guiding techniques to also navigate the user towards the target. A well-known technique including navigation was suggested by Biocca et al. [5] giving path-like information. Their Omnidirectional Attention Funnel could reduce search times, task load and the error rate in comparison to audio cueing and highlighting. For guiding the user to picking targets in a shelf, Schwerdtfeger and Klinker [28] evaluated two similar approaches how to signal the orientation of the target to the user. They compared a tunnel of frames leading from the user's view towards the target on a bezier spline with an in-view arrow pointing towards the target, extended by a rubber band. They found that while participants could orient faster and more accurate when seeing the arrow, they could benefit from the distance information of the tunnel. One has to mention, however, that there were some issues with correctly visualizing the tunnel on the bezier spline at this stage of research. In the same line of approaches, Reif and Günthner [21, 22] compared AR-based picking with a pick-by-light system. They present an in-situ AR guiding showing a tunnel of circles on a curve from the user's view towards the target, similar to the Attention Funnel. The test environment comprised eight shelves with 6 aisles in a warehouse. They could show that participants were faster and made fewer errors using AR-based guiding. Additionally, the authors report a high acceptance. Hanson et al. [11] tested a similar ring-based picking assistance against classical paper instructions both for single kit preparation and for batch preparation. Their picking guidance is a tunnel of rings from the user's FOV towards the target. For batch preparation, additionally the amount of items to pick is shown in-situ. We suggested a spline-based path (a detailed description can be found in 3) for picking and placing in environments where targets can be possibly occluded [23], explicitly giving a route to follow. It was shown to be faster than highlights and the SWave technique.

Several conventional techniques are well-established in industry to support order picking. In order to measure if AR-based techniques can compete with these, Thomas et al. [30] conducted a study in an environment consisting of two shelves and a picking cart. The conventional techniques they tested are 1. pick-by-paper (without any verification that the correct object was picked), 2. pick-by-paper with barcode-scan verification and 3. pick-by-light with button-press verification. They state that all these techniques are popular approaches in industrial order-picking. As an AR-based method, they tested pick-by-HUD using Google Glass with verification by RFID scan. A wearable RFID scanner is used for this purpose. The AR-based approach led to fastest and most accurate results in their evaluation. Moreover, this approach was preferred by most participants. Latif and Shin [17] developed a path-based guiding approach for picking tasks similar to the approach previously presented [23]. They also apply an A* algorithm to find a path in a pre-scanned environment. In contrast to the refined approach presented here, the walkable areas are not automatically processed during runtime, but pre-processed in the Blender modelling software. The authors found that study participants were 23,6% faster than using a paper-based system. An approach combining different techniques was proposed by Puljiz et al. [20] for workers in an autonomous warehouse. They combined a path for navigating through the aisles and an arrow for maneuvering. Additionally, a minimap can be opened to see the live status of the warehouse including the positions of robots. Their system however was in early stage at the time of publication, thus the authors did not evaluate it.

Given that AR-based guiding is helpful in principle, the question arises if this is also applicable in a full-shift, where a worker is wearing smart glasses for 8 hours. Murauer and Planz [18] tackled this question in a warehouse scenario. They compared textual picking instructions on binocular AR glasses and on a cart-mounted monitor. Their results show a decrease in task-completion-time and error-rate using the AR glasses. They observed that the performance measurements improved over the day. However, a Simulator Sickness Questionnaire indicated a higher disorientation at the end of the shift when using AR glasses. The authors report that still no participant opted out during the shift due to discomfort. Elbert and Sarnow [6] aimed to find out how to employ AR in picking processes respecting cognitive ergonomics. Besides a literature review, they conducted qualitative interviews with 11 pickers. They found that font size (if text is necessary to convey all required information) should be adjustable for the individual user. They propose a support level for novice which only highlights the target items, experienced pickers could also be able to operate a quantity control. A static list (e.g. on a HUD) is not seen as more useful than standard screens which are already available in picking.

2.3 Considerations towards a taxonomy of attention guiding techniques

Guiding the users attention in an efficient way is not a trivial task. As the approaches and evaluations presented in the last sections show, different challenges have to be faced using AR-based guiding depending on the task and the work environment. These define the requirements an attention guiding technique has to meet: In workbench applications, often only the position of one or more targets on the desk in front of the user is relevant. All possible targets are visible for the user. In a workshop environment, however, targets can be located in shelves next to or behind the working place. Then, also a re-orientation of the user is necessary. Sometimes, only showing where to orient to can be sufficient, sometimes, the actual position has to be guided to. In warehouses, picking objects often requires to navigate through multiple aisles. Then, it can also be relevant if the guiding information can handle occlusions. For different areas of application, specific hardware has to be chosen. This choice can be constrained by ergonomic factors or the price of different devices. Nearly everyone owns a smartphone and knows how to handle it. Buying a Microsoft HoloLens can open up new possibilities, but is on the other hand expensive and may restrain a workers sight. Finally, the effort to implement an attention guiding technique can differ significantly. If, e.g., a path planning algorithm is used, all obstacles have to be registered with it.

This can be condensed into five factors determining the choice of an adequate attention guiding technique based on the requirements of the task and work environment:

Degrees of Freedom
Depending especially on the size of the environment, a technique can range from exclusively giving orientational cues (on two or three axes) or positional cues, to combinations including information about the distance or the path to take.

Supported Phases
The process of finding a target using attention guiding can be divided in up to three *phases*: If the target is not visible initially, the user orients towards the target direction. If the user is not close to the target, he navigates to its rough location. Finally, he maneuvers to the exact position, e.g. in a shelf.

Table 1: Classification of Attention Guiding Techniques

Technique	DoF	Supported Phases	Flexibility	Integration	Investment
Highlight	3D pos	Maneuvering	Multiple targets	in-situ & occlusions	see-through hw
Minimap	2D pos	[Navigation]	Multiple targets	HUD	all mobile devices
Arrow	2D/3D rot	Orientation	One target	HUD	all mobile devices
SWave	3D pos, 3D rot	Orientation, Maneuvering	One target	in-view & in-situ, no occlusions	3D glasses
Line	3D pos, 3D rot	Orientation, Maneuvering	One target	in-view & in-situ, no occlusions	see-through hw
Attention Funnel	3D pos, 3D rot	Orientation, Maneuvering	One target	in-view & in-situ, no occlusions	3D glasses
3D Path	3D pos, 3D rot	All	One target	in-view & in-situ, occlusions	3D glasses, spatial understanding

Flexibility
Guiding techniques can be able to handle *multiple targets* at the same time or only one.

Integration
An attention guiding technique can be presented in an abstract way on a HUD, e.g. showing a compass, or it can integrate with the environment giving in-situ information. Some techniques need to handle occlusions for optimal integration, especially if the navigation phase is supported, the user has to be guided around obstacles which occlude the target.

Investment
Also the *hardware* on which the guiding technique plays a role. Some techniques can be executed on any mobile display, e.g. a smartphone. In-situ techniques require see-through hardware which can be video-see-through on smartphones or see-through HMDs. For conveying depth information, 3D capabilities are needed and therefore such techniques can only be used with binocular see-through HMDs. Finally, if obstacles and occlusions have to be computing during runtime for computing a path, spatial understanding is required which is supported by mixed-reality glasses like Microsoft HoloLens.

In table 1 the different attention guiding approaches from the literature are classified regarding the five factors. The classification reveals that all techniques have specific advantages and disadvantages. E.g., a highlight is only helpful if in the FOV of the used AR device, but it is capable of visualizing multiple targets at the same time. Showing a compass-like arrow allows using cheap mobile devices, but not fine-grained position can be displayed. Guiding using a path provides the user with all necessary information to find a target, but it usually requires expensive AR hardware.

Our aim is to evaluate these different categories of guiding techniques for large-scale scenarios like warehouses, not only regarding their performance, but also if they are suitable for usage in work environments. The aim is to cover a relatively broad area of the factors presented before and moreover especially take into regard the ergonomics of the techniques. Another focus is especially the autonomy they leave to the users.

3 ATTENTION GUIDING TECHNIQUES FOR LARGE ENVIRONMENTS

In large environments, targets will usually be located in areas which are not directly visible for the user as they are occluded by other objects, like book shelves in our case. The user has to navigate around obstacles to reach the target.

Intuitively, one would suppose that techniques which support the required navigation phase should outperform other techniques. The user should be guided towards the target on the shortest way. However, people might feel patronized by the technique and might prefer to plan their path by themselves. If users are looking for more than one target, they also might prefer to choose the order of reaching these targets by themselves. In general, this is a question of autonomy: Do people prefer to plan themselves or do they they prefer to just follow a pre-planned path. This is especially important for productive use when users are supposed to use the assistive system for a long period, e.g. in order picking.

To find out about techniques leading to fastest picking times, but also considering the open questions explained before, six attention guiding techniques were designed, reflecting these aspects. The choice of techniques is based on the literature presented in this paper as well as on findings of previous studies [23, 25, 26] and shall cover different aspects of attention guiding factors (see section 2.3). In the following, these techniques are explained in detail. Example screenshots of them can be found in Figure 1. In table 1, these techniques are classified in terms of the proposed taxonomy.

3.1 Highlighting Techniques

The first two techniques do not support the different phases of attention guiding but simply highlight the target location or locations. We chose highlights which are not occluded by the environment to make sure that people can directly use the guidance.

X-Ray Highlight. An in-situ arrow (Figure 1 a) is placed right on top of the target object, pointing towards it. Like using an x-ray, this arrow is not occluded by other objects in the way, thus always visible for the user. This can basically regarded as a highlight, which is not occluded. In this stage, the user would have to look around until finding the arrow. This would inherently cost some time in comparison to techniques which support the orientation or even navigation phase. To prevent this, a small radar-like arrow is displayed at the edge of the display pointing towards the highlight.

Multi X-Ray Highlight. The multi x-ray highlight (Figure 1 b) is a variant of the guiding technique described before. While the normal X-Ray highlight only shows one target at the same time, the Multi X-Ray highlight supports multiple targets. For each highlight which is not in sight, the small radar-like arrow is displayed.

3.2 Path-based Techniques

Path-based techniques support all phases of attention guiding. For a single target, the user simply has to follow the displayed path. For multiple targets, many paths at the same time could clutter the field-of-view, thus more sophisticated solutions are required.

3D Path. In contrast to the highlight techniques described before which only support the orientation phase and apart from that giving positional information, the 3D path (Figure 1 c) also supports the navigation phase. The 3D path always starts in the lower field-of-view of the user, thus at least the first part of it is always visible for the user. Using an A* algorithm on a grid of automatically computed waypoints, control points for the path are identified. The path is then constructed from these points using a Catmull-Rom spline.

The 3D path was initially implemented and evaluated in Virtual Reality [23]. For the present evaluation, several enhancements were implemented: Before, the waypoints for the A* algorithm were manually set, here they are dynamically created. The basis for this is the spatial mapping capability of the Microsoft HoloLens. It automatically uses SLAM to create a mesh-like representation of the environment. From the mesh, the spatial understanding capability can create a grid of points that are detected as floor. The maximum precision is a grid of 8x8 cm. For saving memory, we reduced this grid size to 20x20 cm. Then, each waypoint is connected to its neighbours. All waypoints up to 0.5 m distance are considered neighbours to prevent the resulting Catmull-Rom spline from having unnecessarily sharp turns. Moreover, after this step each waypoint is rated due to its number of neighbours. The A* algorithm is altered to prefer higher rated waypoints. The result is that generated paths are less likely to be close to obstacles and stay in the middle of two close obstacles.

Some minor enhancements encompass guaranteeing the visibility of the beginning of the path also if it starts with a 180° turn as well as smoothing the visualized path when the user is walking.

MultiPath. With a newly developed extension of the 3D path it is possible to support multiple targets. To prevent from cluttering the user's field-of-view, these cannot start directly in front of the user. Instead, each path ends close to the first position which is visible for the user, as can be seen in Figure 1 d. The visible end of the path is highlighted by a sphere. Radar-like arrows point towards the spheres to let the user know in which direction he has to start. When the user comes closer and the next part of the path is not occluded anymore, the path "flies" into the direction of the target until again only a small part of it is visible (or the target itself is visible). The movement can be compared to a balloon which is pulled towards the target.

Technically, the MultiPath creates a single 3D path for each target. In contrast to normal path generation, from the end of the path the first visible waypoint is searched for. All waypoints from the beginning of the path up to this waypoint are discarded. This way, the shortest path which is partly visible for the user is shown. In order to highlight the visible end of the path, the last waypoint is set 30cm higher. A Sphere is attached to the end for additional visibility, letting the path look like a balloon which is bound to the target by a string. When the user moves, the path is recalculated. If the has to be shortened because another waypoint is the first one visible, the path is interpolated over time such that it looks like as if someone was pulling the balloon towards the target. If the path becomes longer (e.g. because the user moves behind an obstacle) the path is not interpolated but directly set.

ArrowPath. The third variant of the 3D path is inspired by head-up-display navigation systems as proposed for cars. Arrows on the floor are visualized instead of a line starting the field-of-view (Figure 1 e). The path ends right in front of the target object.

The ArrowPath is realized using the 3D path algorithm. The path is projected on the floor and points of the Catmull-Rom spline are sampled to show arrows in equal distances. The arrows point in the direction of the tangent of the sampled point on the curve.

3.3 Minimap

Inspired by computer games, many users are familiar with an interactive map of the environment. The Minimap (Figure 1e) is visualized a bit lower than the user's field-of-view, thus it is visible when looking slightly down. A cone originating from a black dot shows the current position and orientation of the user in the environment. Static obstacles are plotted in black, target areas in light gray. The actual targets are highlighted using arrows.

3.4 Hypotheses

The six attention guiding techniques taken into consideration can be split into highlighting techniques, path-based techniques and the Minimap as an overview technique.

Highlighting techniques mainly give positional information and do not support the navigation phase. In our implementation, the additional radar-like arrows in the FOV provide a minimal support for the orientation phase. Using the x-ray variant, occlusions by obstacles are not handled on purpose, allowing users to always see the target. The three path-based techniques support the navigation phase and by design also the orientation and maneuvering phase in the case of 3D Path and MultiPath. The ArrowPath is not always in-view as the arrows stick to the floor, thus orientation and maneuvering are not supported at the same level as by the other path-based techniques. All approaches handle occlusions. The Minimap gives an overview over the environment and all targets. Finding targets requires mapping from the Minimap to the real environment. In exchange, the user's view is completely free as long as he decides to actively look at the map.

For rating attention guiding techniques, we take three different categories into consideration: Firstly, we consider the task-completion-time as well as the distance they move to get there as the performance of solving a task. Secondly, we suppose that the techniques will lead to a different feeling of autonomy. Finally, we also take ergonomics and task load into consideration as users might wear similar assistance systems for several hours in the future.

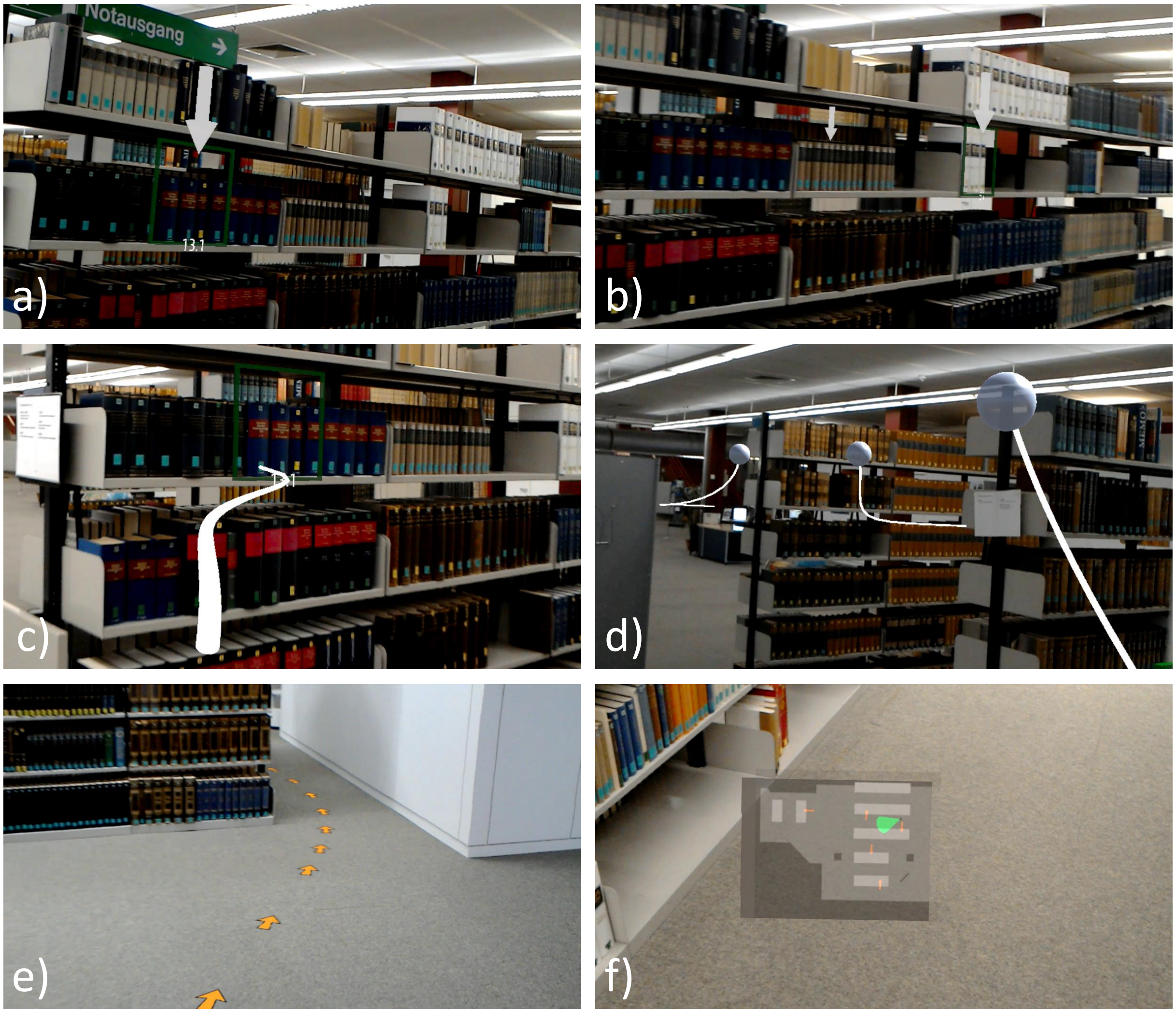

Figure 1: Attention guiding techniques for large-scale environments: a) The single X-Ray highlight, b) the Multi X-Ray highlight for multiple targets, c) the 3D Path showing the way to a target and always at least partly visible, d) the MultiPath showing paths to multiple targets, e) the Arrow Path showing a path on the floor and f) the Minimap which is visible if the user looks a bit downwards.

Performance. As path-based techniques support all phases of attention guiding, these should outperform highlighting techniques. The user simply has to follow the *optimal path* towards the target. As the Arrow Path provides a less direct guiding, user will need more time to initially see the path and finally find the fine-grained target position. For multiple targets, the user can choose between different paths, which probably leads to a non-optimal picking order. Therefore, we suppose a decrease in performance. The Minimap technique is supposed to lead to slowest task-completion-times, as the user has to completely plan the path to each target.

Autonomy. We assume the multi-target techniques to increase the user's feeling of autonomy as they inherently leave more decisions to the user. The same goes for the highlighting techniques which support less phases of attention guiding and thus leave more work to the user. A special case is the Minimap which can be used to pre-plan the task completely. This should arise a very high feeling of autonomy.

The 3D Path technique on the other hand generates visualizations which completely define the user's movement. The feeling of autonomy should be lowest.

Ergonomics. We assume ergonomics to be best for techniques which do not lead to unnatural movements. Projecting guiding information on the floor, the ArrowPath requires the user to look downwards regularly. Similarly, the user has to look downwards to display the Minimap. The 3D Path should require least head movements: Ideally the user does not have to look around at all to find a target, but is directly steered towards it. The other way around, the single X-Ray Highlight is assumed to generate most head movements to find the targets.

The task load should be negatively correlated with the degrees of freedom of a technique and the supported phases: The Minimap gives few information which should generate a high task load. The other way around, users are expected to report lowest task load when using the 3D Path.

4 EVALUATION OF GUIDING TECHNIQUES

We chose our university library for comparing the different guiding techniques. The book shelves of a library are very similar to shelves in an industrial warehouse regarding the task of picking objects. In this scenario, we conducted a within-subject experiment. The six guiding techniques as described in Section 3 were evaluated as independent variable. The order of evaluated techniques per subject was determined by a latin-square design.

As objective data, we measured the time until participants reached a target book, the distance they walked to get there and the angle they moved their heads. For each target, we also logged the shortest path computed by the A* algorithm of the path-based techniques. Additionally, we measured the share of time that participants had an assistive visualization in the field-of-view of the glasses.

As subjective data, after each 3 repetitions of the task using one guiding technique, participants were asked to fill out a questionnaire about their rating of the technique in the terms of speed, correctness, ease-of-use, occlusion of obstacles, usefulness and fun. Moreover, they were asked to report about their task load using the NASA TLX questionnaire. At the beginning of an experiment, participants additionally were asked about demographic data as well as about their familiarity with virtual reality and computer games. Also they reported about their ability to navigate in unknown environments. After the experiment, participants were also asked about their preferred guiding technique and if they preferred multi- or single-target techniques.

30 Participants took part in the experiment. 11 of them were male, 19 female. Most participants were students of our university from various fields of study, their average age was 26.9 years (sd: 4.2). Only six of them had reasonable experience with virtual reality, eight had little experience. 16 of them were highly experienced in computer games.

4.1 Scenario

The study took place in a section of our university library. We chose an area which was usually unoccupied by students, thus it was possible to conduct the experiment during the regular opening times of the library. The area comprised six shelves: Four in a row and two more orthogonal to them. Each shelf had two sides with books, so effectively one can see them as 12 shelves. All in all, the walkable space was approximately $100m^2$.

The task for participants was to find positions of books in the shelves. For each guiding technique, we asked participants to do three repetitions of finding five books. In each shelf, six locations for books (three on each side) were evenly distributed. Thus, in total 36 books could be chosen for picking. In each repetition of 5 books to find, one book each in five of the six shelves was randomly chosen. This procedure ensured that the total path to walk was similar for each repetition.

4.2 Presentation and Interaction

The experiment was conducted with a Microsoft HoloLens which has a diagonal field-of-view of about 35°. No additional hardware (e.g. a laptop) was used, everything was done in real time on the HoloLens. Using its inside-out tracking capabilities, the library environment did not have to be equipped with external sensors. The spatial mapping and spatial understanding capabilities of the MixedRealityToolkit were used for generating waypoints for calculating paths. For interaction, the gesture detection of the HoloLens was used: Participants were asked to perform a "tap-gesture" to confirm they found a book. Additionally, a bluetooth keyboard served as input device for setting up the experiment by the operator as well as for starting the next repetition by the participants.

4.3 Procedure

To make sure that the guiding techniques were exactly identical for each participant, after an initial scan of the library environment, the technical spatial mapping and understanding process of the HoloLens was done once during preparation of the study. The generated waypoints were saved and kept constant for the whole study. The same goes for the target books which locations were distributed as described before and then annotated by hand.

Before each experiment, the library environment was re-scanned to make sure that the tracking of the HoloLens worked flawlessly. After filling out the initial questionnaire, we explained the different guiding techniques and how to setup and adapt the HoloLens. Then the device was calibrated for the participant. After a short introduction in how to conduct the "tap-gesture", the experiment started. As described before, between changing the guiding techniques two questionnaires were filled out and finally participants were asked about their preferences. Then, they were refunded for taking part in the study (which took them approximately 60 minutes) with 10€.

5 RESULTS

Firstly, the objective data which were recorded will be presented. After that, the subjective questionnaire results are shown.

5.1 Objective Results

The 30 participants were guided to 2700 target positions. On average, participants needed 13.4s (sd: 6.05) to reach a target book. They walked a distance of 9.9m (sd: 6) per target and thereby turned their head 569.6° (sd: 303.5). As the objective results for time-on-task, distance and head angle were skewed and thus violated the normality assumption for an ANOVA, the Aligned Rank Transform for non-parametric analysis [33] was used to preprocess the data. The post-hoc analyses were then done using pairwise Tukey-corrected Least Squares Means.

The MultiPath guiding technique required least time on average to pick one book, with a mean of 12.58s per target (sd: 5.91s) followed by the Multi X-Ray Highlight with 12.88s (sd: 6.03s). Using the X-Ray Highlight participants needed 13.06s (sd: 5.78s), with the 3D Path guidance it was 13.13s (sd: 5.94s). The technique taking most time to find a target was the ArrowPath with 14.62s (sd: 6.0s).

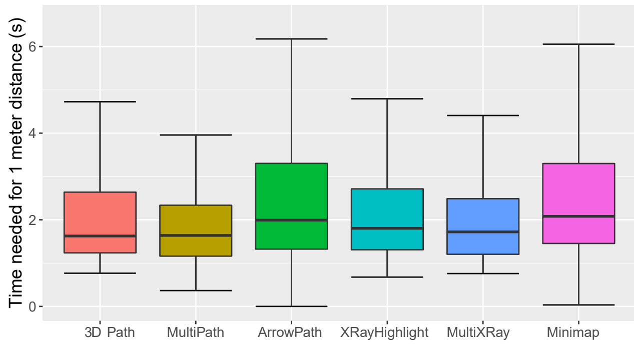

Figure 2: The time participants needed to find a target book in the library area, normalized with the minimal distance they had to walk from the start to the target.

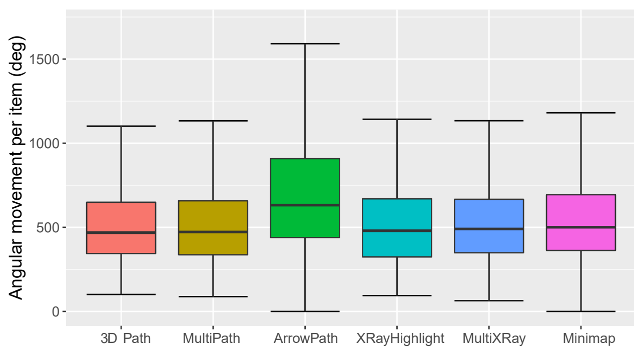

Figure 3: The angle participants looked around and changed their orientation to find a target book in the library area.

The ANOVA revealed a significant difference between the techniques (p<.001). The post-hoc tests showed that all techniques except the Minimap led to faster task-completion-times than the ArrowPath guidance (3D Path p=.0015, MultiPath p<.001, X-Ray Highlight p=.0015, Multi X-Ray Highlight p<.001). Moreover, the Multi X-Ray Highlight was significantly faster than the Minimap technique (p=.0136).

Figure 2 shows the time to find a book normalized by the minimal distance participants had to walk to get there. Also here, the ANOVA showed a significant difference (p<.001). In addition to the significant differences of the analysis of the raw times, the 3D Path and the Multi X-Ray Highlight were significantly faster than the Minimap (p<.001). Moreover, the Multi X-Ray Path was significantly faster than X-Ray Highlight (p=.0149).

The distances participants walked to reach a target did not show significant differences in the ANOVA.

The average angular head movement participants made when being guided to a target (see Figure 3 was similar for all techniques except the ArrowPath (3D Path: 532°, sd: 280°; MultiPath: 529°, sd: 281°; X-Ray Highlight: 549°, sd 309°; Multi X-Ray Highlight: 546°, sd :287°; Minimap: 561°, sd: 285°). The ArrowPath technique showed an average head movement of 704° (sd: 342°). The ANOVA showed significant differences for the angular head movement (p<.001). The post-hoc tests revealed a significantly higher head movement for the ArrowPath (for all comparisons p<.001).

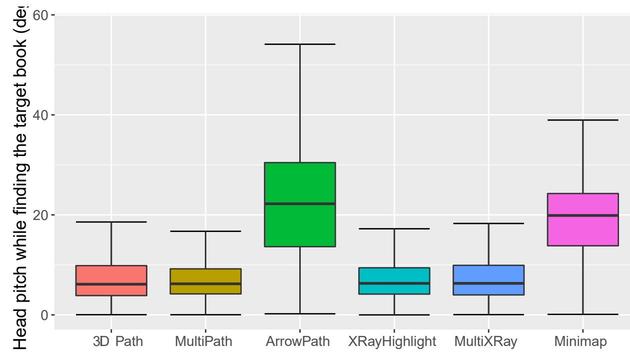

Figure 4: The head pitch of participants being guided towards target books.

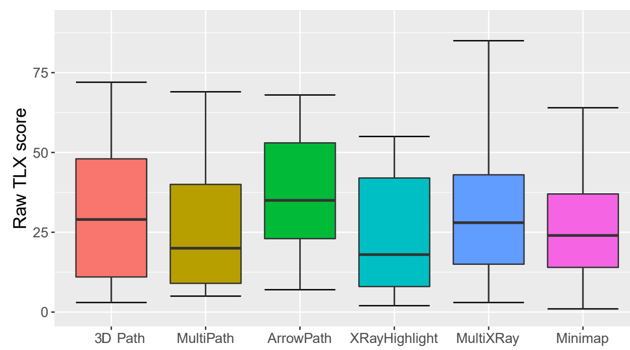

Figure 5: The results from the NASA TLX questionaire.

For analyzing the ergonomics of the techniques, a special focus was on the head pitch (Figure 4). The ArrowPath and Minimap techniques led participants to look more towards the floor: The mean head pitch for the ArrowPath guidance was 41.3° (sd: 77.9°), for the Minimap it was 32.2° (sd: 56.1°). Using the other techniques, participants mainly looked straight. With the 3D Path, the mean head pitch was 12.8° (sd: 30.6°), 9.8° (sd: 17.7°) for the MultiPath, 10.7° (sd: 20.3°) for the X-Ray Highlight and 9.5° (sd: 14.8°) for the Multi X-Ray Highlight. The ANOVA showed significant differences between the techniques (p<.001). The post-hoc tests revealed that the head pitch of both the ArrowPath and the Minimap guidances was significantly higher than for all other techniques (p<.001). There was no significant difference between the two techniques.

Additionally, the time a guidance was in view was measured. Inherently, the 3D Path is always visible in the field-of-view of the HoloLens. The MultiPath was not necessarily in view, it was looked at in 83.3% (sd: 21%) of the time participants looked for target books. The highlighting techniques were in view more seldomly: The X-Ray highlight in 63.9% (sd: 25.9%) of the time and the Multi X-Ray highlight in 70.4% (sd: 25.5%). The technique least often in view was the Minimap with 50.5% (sd: 30.2%). The ANOVA and the post-hoc tests revealed that all differences are highly significant (p<.001).

5.2 Subjective Results

The results of the NASA TLX questionnaire and the preference questionnaire, both filled out by participants after using each guiding technique, can be found in Figure 5 and Figure 6.

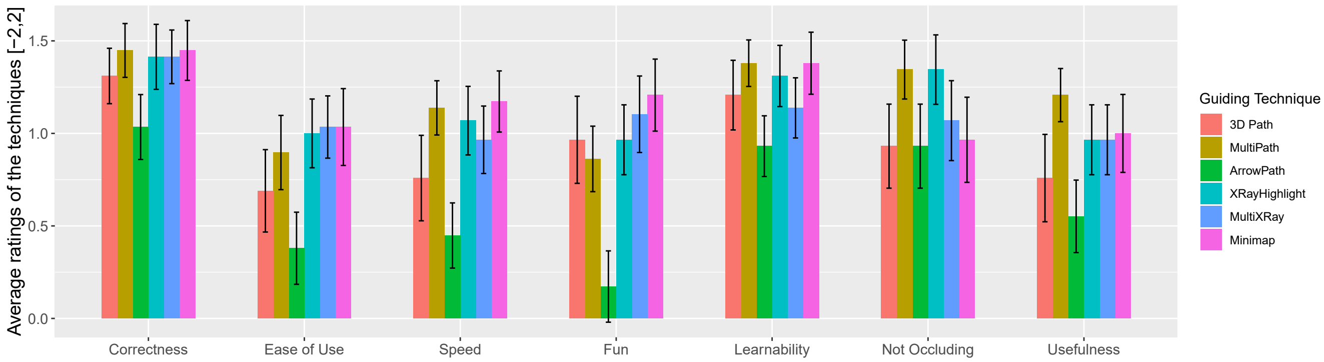

Figure 6: The subjective ratings of the guiding techniques.

Participants reported the highest task load using the ArrowPath with a raw TLX score of 38.4 (sd: 21.7), followed by the 3D Path with a score of 31.8 (sd: 22.8). For the Minimap, a score of 30.8 (sd: 23.2) was reported, 29.4 (sd: 19.1) for the Multi X-Ray Highlight and 27.3 (sd: 20.4) for the MultiPath. The lowest task load was reported for the X-Ray Highlight with a score of 24.8 (sd: 22.3). The ANOVA did not reveal any significant differences.

In the preference questionnaire, guiding techniques were rated on a five-level Likert scale (transferred to the range [-2,2]) in terms of correctness, ease of use, speed, fun, learnability, usefulness and occlusion of real world objects. All results can be found in Figure 6. In general, participants rated the techniques rather positively. In terms of correctness, ease of use, speed and fun, the Minimap technique got the highest ratings, followed by the MultiPath which was moreover rated best in terms of usefulness and learnability. In all these categories, the ArrowPath was rated worst. The MultiPath and the X-Ray Highlight were rated as occluding the environment least, followed by the Multi X-Ray Highlight. However, all techniques were rated as only little occluding.

As an answer to the final question which guiding technique could guide them best, 9 participants chose the MultiPath, 7 the Multi X-Ray Highlight, 5 the 3D Path, 4 the Minimap and each 1 the ArrowPath and the X-Ray Highlight. 20 of the 30 participants stated they would prefer being guided by a technique supporting multiple targets at the same time instead of going to one after another.

6 DISCUSSION

In contrast to our expectations, the 3D Path was not significantly faster than the other guiding techniques, even though it is supposed to lead the user to the target on the shortest path. Indeed, on average the MultiPath technique even was slightly faster. Also performance using the Multi X-Ray Highlight was slightly faster than with the regular X-Ray Highlight which showed target books in the order of the shortest path. We assume participants could use the knowledge about all targets to pre-plan where to go next.

The distance participants walked to a target did not significantly differ with all guiding techniques. Thus, on the one hand no technique misled participants, on the other hand this suggests that participants were able to plan their way efficiently. The angular head movement is significantly higher when participants used the ArrowPath technique. Here, the user has to look for the arrows on the floor. The other techniques all at least have hints to let the user know where to look for guidance next. This shows that the hints worked for all techniques and ensured that participants did not have to search the environment.

The ArrowPath guiding technique in general showed a weak performance and was rated accordingly. One reason can be supposed to be the missing guidance towards the path on the floor. Moreover, participants had to look down to see the path, which might have resulted in a more unstable tracking of the HoloLens.

In contrast, the Minimap technique also showed rather weak objective results, but it was rated very positively by the participants. They had a high feeling of autonomy, which could be expected as this technique is the only one where users can pre-plan their path completely. Especially participants with experience in computer games enjoyed using the Minimap, a number of them reported that it reminded them of similar maps in games. Interestingly, participants estimated their performance regarding time-on-task as the fastest which does not match the measured time. As the Minimap was also rated to have the highest learnability, maybe there is a steep learning curve which could lead to faster performance after more repetitions. As hypothesized, both the Minimap and the ArrowPath technique made participants look down significantly more than using the other techniques. Thus, considering ergonomics, these techniques are not optimal.

In general, two third of the participants preferred guiding techniques which support multiple targets. This leaves more autonomy to the user. Actually, as the MultiPath technique had the best performance, one could suppose that users can use this autonomy to pre-plan their paths better. Looking at the task load participants reported, it seems that planning by themselves did not increase cognitive effort significantly. However, as expected, the lowest task load was reported for the basic X-Ray Highlight technique.

7 CONCLUSION

In this paper, we evaluated several AR-based attention guiding techniques with a focus on usage in large-scale environments like warehouses. Starting with an overview over the different aspects and factors of attention guiding techniques, we classified common techniques and selected appropriate ones for evaluation in a large-scale picking environment. Based on the insights from the literature, we moreover proposed a novel MultiPath attention guiding technique, which can be seen as a compromise between direct path-based guiding to one target and ways of highlighting different targets at the same time.

The literature analysis, the consequent classification of attention guiding techniques as well as the evaluation in a large-scale environment give several insights for selecting an appropriate guidance for a specific scenario. Main recommendations for selecting a technique can be condensed as:

- Orientational cues are required to make sure users quickly find the correct direction where to go.
- Users benefit from information about other targets such that they are as fast or even faster when showing multiple targets (combined with path-based information) in contrast to showing the best way to each target one after another.
- Users tend to prefer guiding techniques which leave some autonomy to them.

Generally, considering the low cognitive load of participants in the study and the results regarding their performance, we can conclude that in-situ attention guiding is usable and beneficial using current hardware like the Microsoft HoloLens.

7.1 Future Work

As this is a piece of work from a series of evaluation, we could collect several insights from applying different attention guiding techniques in multiple environments. As a final step, we are going to assemble a system which automatically integrates different techniques and activates the best one for each situation.

REFERENCES

[1] H. Alvarez, Iker Aguinaga, and Diego Borro. 2011. Providing guidance for maintenance operations using automatic markerless Augmented Reality system. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 181–190.

[2] Patrick Baudisch and Ruth Rosenholtz. 2003. Halo: A Technique for Visualizing Off-screen Objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 481–488.

[3] B. Besbes, S.N. Collette, M. Tamaazousti, S. Bourgeois, and V. Gay-Bellile. 2012. An interactive Augmented Reality system: A prototype for industrial maintenance training applications. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 269–270.

[4] Sidharth Bhatia and P. Vijayakumar. 2012. Augmented Reality Repair Augmented Guidance System. *International Journal of Computer Science Issues (IJCSI)* 9, 4 (2012), 333.

[5] Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. 2006. Attention Funnel: Omnidirectional 3D Cursor for Mobile Augmented Reality Platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1115–1122.

[6] Ralf Elbert and Tessa Sarnow. 2019. Augmented Reality in Order PickingâĂŤBoon and Bane of Information (Over-) Availability. In *International Conference on Intelligent Human Systems Integration*. Springer, 400–406.

[7] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. 1997. A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. *Personal Technologies* 1, 4 (1997), 208–217.

[8] Steven Feiner, Blair Macintyre, and DorÃľe Seligmann. 1993. Knowledge-based Augmented Reality. *Commun. ACM* 36, 7 (July 1993), 53–62.

[9] Markus Funk, Andreas Bächler, Liane Bächler, Oliver Korn, Christoph Krieger, Thomas Heidenreich, and Albrecht Schmidt. 2015. 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)*. ACM, New York, NY, USA, 1:1–1:8.

[10] Ankit Gupta, Dieter Fox, Brian Curless, and Michael Cohen. 2012. DuploTrack: A Real-time System for Authoring and Guiding Duplo Block Assembly. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 389–402.

[11] Robin Hanson, William Falkenström, and Mikael Miettinen. 2017. Augmented reality as a means of conveying picking information in kit preparation for mixed-model assembly. *Computers & Industrial Engineering* 113 (2017), 570–575.

[12] S. Henderson and S. Feiner. 2011. Exploring the Benefits of Augmented Reality Documentation for Maintenance and Repair. *IEEE Transactions on Visualization and Computer Graphics* 17, 10 (2011), 1355–1368.

[13] S. J. Henderson and S. Feiner. 2009. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. 135–144.

[14] Klaus Kammerer, Rüdiger Pryss, Kevin Sommer, and Manfred Reichert. 2018. Towards Context-Aware Process Guidance in Cyber-Physical Systems with Augmented Reality. In *2018 4th International Workshop on Requirements Engineering for Self-Adaptive, Collaborative, and Cyber Physical Systems*. IEEE, 44–51.

[15] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. La Viola, T. Mashita, and H. Takemura. 2014. The Effectiveness of an AR-based Context-Aware Assembly Support System in Object Assembly. In *2014 IEEE Virtual Reality (VR)*. 57–62.

[16] Robert E. Kraut, Susan R. Fussell, and Jane Siegel. 2003. Visual Information As a Conversational Resource in Collaborative Physical Tasks. *Hum.-Comput. Interact.* 18, 1 (June 2003), 13–49.

[17] Ummi Khaira Latif and Soo Young Shin. 2019. OP-MR: the implementation of order picking based on mixed reality in a smart warehouse. *The Visual Computer* (2019), 1–10.

[18] Nela Murauer and Nerina Pflanz. 2018. A full shift field study to evaluate user-and process-oriented aspects of smart glasses in automotive order picking processes. *IxD&A* 38 (2018), 64–82.

[19] N. Petersen, A. Pagani, and D. Stricker. 2013. Real-time modeling and tracking manual workflows from first-person vision. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 117–124.

[20] David Puljiz, Gleb Gorbachev, and Björn Hein. 2018. Implementation of Augmented Reality in Autonomous Warehouses: Challenges and Opportunities. *arXiv:1806.00324 [cs]* (June 2018).

[21] Rupert Reif, Willibald A. Günthner, Björn Schwerdtfeger, and Gudrun Klinker. 2009. Pick-by-vision comes on age: evaluation of an augmented reality supported picking system in a real storage environment. In *Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*. ACM, 23–31.

[22] Rupert Reif, Willibald A. Günthner, Björn Schwerdtfeger, and Gudrun Klinker. 2010. Evaluation of an augmented reality supported picking system under practical conditions. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 2–12.

[23] Patrick Renner, Jonas Blattgerste, and Thies Pfeiffer. 2018. A Path-based Attention Guiding Technique for Assembly Environments with Target Occlusions. *IEEE Virtual Reality 2018* (2018).

[24] Patrick Renner and Thies Pfeiffer. 2017. Attention Guiding Techniques using Peripheral Vision and Eye Tracking for Feedback in Augmented-Reality-based Assistance Systems. In *3D User Interfaces (3DUI), 2017 IEEE Symposium on*. IEEE, 186–194.

[25] Patrick Renner and Thies Pfeiffer. 2017. Augmented Reality Assistance in the Central Field-of-View Outperforms Peripheral Displays for Order Picking: Results from a Virtual Reality Simulation Study. *ISMAR 2017* (2017).

[26] Patrick Renner and Thies Pfeiffer. 2017. 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, 89–92.

[27] Torben Schinke, Niels Henze, and Susanne Boll. 2010. 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 (MobileHCI '10)*. ACM, New York, NY, USA, 313–316.

[28] Bjorn Schwerdtfeger and Gudrun Klinker. 2008. Supporting Order Picking with Augmented Reality. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '08)*. IEEE Computer Society, Washington, DC, USA, 91–94.

[29] D. Stanimirovic, N. Damasky, S. Webel, D. Koriath, A. Spillner, and D. Kurz. 2014. [Poster] A Mobile Augmented reality system to assist auto mechanics. In *2014 IEEE International Symposium on Mixed and Augmented Reality)*. 305–306.

[30] Charu Thomas, Theodore Panagiotopoulos, Pramod Kotipalli, Malcolm Haynes, and Thad Starner. 2018. RF-pick: comparing order picking using a HUD with wearable RFID verification to traditional pick methods. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. ACM, 168–175.

[31] Marcus Tonnis and Gudrun Klinker. 2006. 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*. IEEE Computer Society, 13–22.

[32] X. Wang, S.K. Ong, and A.Y.C. Nee. 2016. Multi-modal augmented-reality assembly guidance based on bare-hand interface. *Advanced Engineering Informatics* 30, 3 (Aug. 2016), 406–421.

[33] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 143–146.