Augmented reality smart glasses for operators in production: Survey of relevant categories for supporting operators

Oscar Danielsson*, Magnus Holm, Anna Syberfeldt

*University of Skövde, PO Box 408, 54128, Skövde, Sweden
* Corresponding author. Tel.: +46-500-448-596. E-mail address: oscar.danielsson@his.se

Abstract

The aim of this paper is to give an overview of the current knowledge and future challenges of augmented reality smart glasses (ARSG) for use by industrial operators. This is accomplished through a survey of the operator perspective of ARSG for industrial application, aiming for faster implementation of ARSG for operators in manufacturing. The survey considers the categories assembly instructions, human factors, design, support, and training from the operator perspective to provide insights for efficient use of ARSG in production. The main findings include a lack of standards in the design of assembly instructions, the field of view of ARSG are limited, and the guidelines for designing instructions focus on presenting context-relevant information and limiting the disturbance of reality. Furthermore, operator task routine is becoming more difficult to achieve and testing has mainly been with non-operator testers and overly simplified tasks. Future challenges identified from the review include: longitudinal user-tests of ARSG, a deeper evaluation of how to distribute the weight of ARSG, further improvement of the sensors and visual recognition to facilitate better interaction, and task complexity is likely to increase.

Keywords: augmented reality; assembly operator; literature survey; augmented reality smart glasses

1. Introduction

Industry 4.0 is one of a number of initiatives that have been undertaken to improve manufacturing, mainly by enabling more customizable production through the use of Information and Communications Technologies (ICT) [1]. However, while technology such as robotics are being used to a greater extent, assembly workers are still likely to have a central role in manufacturing operations [2]. An increased need for flexibility and adaptability in future production systems is likely to lead to a demand for cognitive aids such as augmented reality (AR) [3]. Production managers and HR managers have previously predicted that support tools on the shop-floor will become increasingly important and several of them mention AR as a probable technology to be integrated [4]. This can now be seen in that while adoption levels of AR are still low in industry in general, there are already examples of AR being used in manufacturing operations [5].

This aim of this paper is to explore the operator perspective of using AR smart glasses (ARSG) in assembly. This will contribute to a better understanding of the current status and future challenges of ARSG in relation to assembly operators and thereby help facilitate a faster application of ARSG in assembly. The paper will achieve this aim by reviewing categories that are relevant for the operator perspective. A previous scoping review of ARSG for industrial assembly operators identified six categories covering an operators perspective: assembly instructions, human factors, design, validation, support, and training (as seen in Figure 1) [6].

The connection between the categories in Figure 1 that was established by [6] can be described as follows: The two main perspectives of ARSG for operators are assembly instructions and human factors. Assembly instructions are the main purpose for operators to use ARSG but human factors is also critical to ensure operator safety. Both of these categories needs to be considered in ARSG design. The design needs to be validated.
and validation in turn depends on how the ARSG are to be used, as a live support in production or as a separate training tool. Based on these connections the categories assembly instructions, human factors, design, support, and training are explored in this paper.

![Operator Perspective Diagram](image)

Fig 1. Operator perspective of ARSG in assembly using categories adopted from [6].

2. Background

There are generally three ways through which a user can experience AR: worn on the user’s head (head-mounted), held in the user’s hand (handheld), or through equipment placed in the user’s environment (spatial) [7, 8]. Handheld solutions are generally unsuitable for operators since they need both hands for assembly tasks. With a spatial solution the operator does not need to wear any extra equipment, but it limits where AR can be displayed to only close to the equipment. It is also limited as it can only display 2D objects on physical surfaces [9].

Head-worn AR can be further categorized into, for instance, contact lenses, helmets, and headsets (smart-glasses) [8]. This paper defines ARSG as a wearable device with one or two screens in front of the user’s eyes that can merge virtual information with physical information in the user’s field of view (FOV). The definition is similar to that used by [10] but broader. The motivation for this is that as ARSG continues to improve it is a reasonable assumption that all head-worn AR will be light and small enough to be considered as smart-glasses. The main advantages of ARSG are that the display is in the operator’s FOV, can display information in full 3D, and is hands-free. The main disadvantages are that ARSG currently have a more limited battery-life and FOV compared to spatial and handheld solutions.

Four ways to implement AR in ARSG is projection based, eye multiplexed, optical see-through, and video see-through [11]. Retinal projection (1 in fig.2) is a fifth way, where thin parallel light beams are focused into the user’s eyes [12]. Projection based AR (2 in fig.2) is implemented with projectors worn on the user’s head and retroreflective materials placed in the environment [13]. Eye multiplexed AR (3 in fig.2) is a virtual scene registered to the physical environment but not composited with the real world view. Video see-through (4 in fig.2) combines virtual content with a real-time video stream of reality and presents the result on a screen in front of the user [14]. Optical see-through (5 in fig.2) creates AR in the user’s FOV, usually by directing the light of the virtual scene through half mirrors or prisms [11]. Optical see-through is currently the most common solution used in commercial ARSG [15]. ARSG displays can be monocular (one eye views a screen, A in fig.2), binocular (both eyes view the same screen, B in fig.2), or dichoptic (each eye views different screens, enabling depth perception, C in fig.2) [16]. Dichoptic is preferable for ARSG if spatially sensitive information should be displayed.

![Different Forms of ARSG](image)

Fig 2. Different forms of ARSG-rendering and display. A. Monocular. B. Binocular. C. Dichoptic. 1. Retinal projection. 2. Projection-based. 3. Eye-multiplexed. 4. Video see-through. 5. Optical see-through.

3. Assembly instructions

Assembly operators need instructions on how to perform his/her assembly tasks, and more instructions are needed the more complex the task is [17]. Since products are updated and replaced regularly, operators need updated instructions to perform the correct assembly. Operators and white-collar workers at three different plants within the same global production network where interviewed by [17] in regards to areas of improvement within assembly instructions. Some problems they identified were slow updating processes (it could take three weeks for instructions to be updated at one plant), a technical language that was hard to understand, irrelevant information, a lack of feedback on errors made, and a large variation in teaching quality due to operators learning from each other. Limits on teaching quality have been identified in other reviews as well [18]. Operators also wanted more individualized and dynamic instructions and which problems that occurred, and their prevalence, varied between the plants [17]. In another case it was found that instructions should focus on clearly marked pictures and be as simple as possible with minimal text [19]. But according to [20] written text should not be removed completely. They found that users using multimedia instructions (both text and pictures) had less errors, faster learning times and were less affected by secondary tasks compared to single media instructions (only text or only picture).

Task complexity also has an influence on how to best design instructions. By dividing users into three experience levels, [21] adapted the instructions to show the right amount of information for each operator. This was implemented in a multi-modal system where the operators used ARSG.

One case study that observed and interviewed operators in an engine assembly factory found no gender or experience differences in how often operators needed to look at assembly instructions [22]. It was further found that the main reasons operators gave for looking at instructions were for checking the torque of the screwing machine, assembly time, and if something goes wrong. In general, the reasons for operators to look at instructions were for things that needed to be checked
equipment used within assembly since it is usually used with a high frequency or for extended periods of time. Ergonomic issues within AR have so far, according to the findings of [23], mostly been tested in laboratory settings within the scientific literature.

An ARSG solution means that some form of equipment will be worn by the operator on his/her head. One important aspect from an ergonomic perspective is the weight of the ARSG. Night vision goggles are another type of head-mounted equipment and [24] found that reducing the length of the protruding part of night vision goggles had little effect on reducing neck muscular strain. The main issue they identified was instead how much weight was placed off from the center of the user’s skull. However, [25] tested different weights and centers of mass for one pair of experimental HMD with different poses. They found that which center of mass (COM) to use varied depending on the pose; if the user was in a neutral position it was best to keep COM around the center of the head, if the user looked up the COM should be placed forward, and if the user looked down the COM should be placed backward, as illustrated in Fig. 3. They also found that a lower mass reduced the neck joint torque ratio, a measure used as an indicator of physical workload. Evaluation of fatigue from extended usage was an aspect identified as valuable future work and [25] further hypothesized that intended duration will determine the recommended upper mass limit due to the strong correlation between duration and load.

![Fig. 3. Shifting of COM depending on head-pose (adopted from [25])](image)

Using a video-based HMD can affect users’ efficiency. When comparing movements and time to finish identification tasks between using and not using an HMD, [26] found that when participants used a HMD to perform a simple object location targeting task they needed more time and made larger movements, implying that using a HMD hinders performance, possibly due to time delays in feedback. The HMD used in the experiment was a form of video-based AR. They also found that the larger movements could affect users’ sickness levels negatively. Areas they identified as interesting for future studies were more extensive studies with more participants and longer exposure time, analyzing simulator sickness and its relationship between posture and performance, as well as if HMDs affect the transfer of training. Similarly, [16] found that video-based HMDs cause significantly more visual discomfort, such as visually induced motion sickness, compared to traditional displays such as TV-screens. Video-based HMD also has an added safety-risk in case of power-failure. Motion sickness in optical see-through HMDs is still an understudied subject according to [27], but they found that participants experienced insignificant motion sickness when using Microsoft Hololens, an optical see-through HMD. This could indicate that an optical see-through HMD would be more suitable in regards to preventing visual motion sickness.

In summary, both the weight and the displacement of the weight of ARSG are important ergonomic factors for operators. The COM should be positioned close to the center of the skull when working in neutral positions, and towards the front or back respectively when looking up or down. Video-based displays can cause significant motion sickness. Microsoft Hololens, an optical see-through HMD, caused insignificant motion sickness which could indicate that optical see-through HMDs cause less motion sickness but further studies are needed.

5. Design

Designing for AR introduces novel challenges and possibilities compared to traditional screen-based interfaces. It is therefore important to know what is known regarding designing for AR in general, and for ARSG in particular.

Designing interfaces for mobile AR requires its own set of design principles compared to general AR and mobile systems in general, so [28] proposed a set of interaction design principles for development of mobile AR applications. The principles were:

1. Use the context for providing content
2. Deliver relevant-to-the-task content.
3. Inform about content privacy.
4. Provide feedback about the infrastructure’s behavior.
5. Support procedural and semantic memory.

The principles were based on mobile AR and the limitations of smartphones but they may still be relevant to ARSG. Using the context for providing content is important since interaction is bound to the physical environment, but this is most important when the physical environment changes. The second principle is to minimize cognitive overhead from interacting with both the system and the real world by minimizing content. Since assembly operators have a high workload this principle is likely to be very relevant. The third principle is probably of lesser relevance in an industrial setting than for private usage, but it
can still be relevant to let operators know what activities are logged. Providing feedback about the infrastructure’s behavior is important since users still interact with real world objects and might depend on external service providers. Applications should therefore be able to adapt to different availability. This principle is of lesser relevance in an industrial setting where all objects the user interacts with can be assumed to be a part of the same infrastructure. The last principle is to support procedural and semantic memory by making the interface and interactions easy to understand which is highly relevant.

A more general set of guidelines, including both AR and VR, and applied to both assembly and maintenance training is proposed by [29]. The first guideline is to start the training with an observation of the task to create a mental model of the assembly. The second is to combine physical and cognitive fidelity since they have complementary advantages. The third is to have the right amount of guidance aids since too much reduces learning. The final guideline is to provide enriched information about the task to promote deep learning. There are however indications that AR will only help an operator if the task is difficult [30].

The operator perspective is also an important aspect of the design. A minimal viable solution for an ARSG-based training system was found based on an engine assembly case [31]. The following features were identified as the most important:

1. The HMD shows the assembly procedure.
2. The HMD shows the relevant parts to pick.
3. The HMD is always available as a training support.
4. The HMD solution works as a “training island” and works separately from the line.

Spatial navigation in an AR interface compared to a traditional screen interface differs in that there is no clear limitation; with a screen a user knows where to look for information but in an AR setting the information could be behind them. A proposed solution to this is a virtual funnel leading the user to the target, a solution that reduced the time needed to find objects and perceived cognitive load for users [32]. This concept has been further explored with different variations such as different forms of the funnel (circular or square) for instance [33]. After six test iterations they arrived at a solution that could guide the user with different visual cues depending on how big the angle was between the user and the intended target. AR might also be used to help operators navigate team tasks by increasing their ambient awareness and by guiding their visual attention [34].

Interaction in the interface will likely differ in an AR implementation compared to a screen-based implementation since the user has a higher degree of mobility and probably do not have a mouse and keyboard in front of them. To make navigation more intuitive, [35] comparatively evaluated a mixed reality (MR) prototype that used a ‘tangible interface’. A physical cube that was tracked by the system allowed the user to navigate in the interface. At the time tracking technology was limited and fiducial markers were used on the cube to allow for it to be accurately tracked. Microsoft Hololens allows for gesture recognition, allowing the user to interact in a similar manner but without an intermediary artifact. Sometimes operators make mistakes and an ARSG system needs to detect these mistakes to allow for correct interaction. Force sensors can detect that parts are picked and placed at the correct position but not that they have the correct orientation, but by combining force sensors with an AR-system more errors can be detected and presented in an ARSG-system [36].

In summary, designing ARSG-interfaces means different challenges compared to a completely digital screen-based interface. AR means placing digital information in the real world and when presented in ARSG this gives the user a hands-free interaction with a bigger environment than a traditional screen-based interface. Design guidelines suggest in general to minimize information in any given context to what is needed in those contexts and to help orient the user to the correct physical location. When interacting in a completely digital world the developer can be seen as omniscient in where all things the user interacts with are. But in AR the world needs to be digitized if the results of interaction are to be interpreted in an ARSG system.

Future challenges lie in improving sensors and visual recognition of parts to allow for more accurate digitizing of the real world. Since ARSG have not been available for a long time or to a wide array of people, guidelines will need to be further tested to ensure their robustness.

6. Support

The role of assembly operators has become increasingly complex, from almost being seen as a machine to now having an increasing number of tasks and responsibilities [4]. Global competition has diversified manufacturing companies’ product range, leading to an increased complexity for assembly workers that in turn affects quality. This can be somewhat alleviated by simplifying the assembly tasks [37]. But due to an increased number of variants and shorter life-cycles of products it is more difficult for assembly operators to achieve task familiarity and routine [38]. While some assembly operator stations currently contain routine work that the operators learn fast there are already stations that require frequent relearning, for instance single inspection point (SIP) stations. Here operators need to inspect different details of products depending on what is currently having quality issues and this can vary from day to day. According to R Lindgren Brewster (personal communication, February 13, 2019), Industrial Business Optimization Manager at Volvo Car Corporation, SIP-stations are complex for operators to learn. The main problem is not to learn new things to inspect, but to stop inspecting things that are no longer a quality issue, leading to waste.

To summarize, some operator tasks are already so complex that learning new tasks, and unlearning old tasks, could benefit from information support through ARSG. Given the shortening of life-cycles of products as well as more simultaneous products it is a likely scenario that task complexity will continue to increase in the future, creating more operator tasks that have a need for increased information support.
7. Training

Training a new operator using on-the-job training (OJT) is one common method of training new operators [39]. Instructions can however be hard to understand for novice operators, who require adequate training before working on the assembly line [40]. This leads to a loss in efficiency that ARSG could help to improve by allowing operators to become independent and efficient workers faster.

AR research for industrial applications has been a research topic since the 1990’s, but there are still severe limitations in that most test-participants are students and assembly tasks are often simplified, many times using LEGO models [41]. AR based training is also mostly compared to paper- or video-based instructions rather than face-to-face training and most measurements are on time rather than quality and training transfer rates [41]. And also, in most studies monitors or handheld devices has been used [41].

Research on training transfer rates from using AR in industrial environments is still very limited. In an effort to close this gap, [42] performed an evaluation of slightly different AR headset interfaces. They found that errors can be reduced by adding a quiz on a task an operator has just been trained on. Most AR training systems are not intelligent but adding intelligent support can significantly improve training results [14]. This seems to support the wish from operators to have dynamic support, found by [17].

In summary, most research regarding AR training for operators has been done using simplified tasks and other equipment than ARSG and it has been done by non-operators outside an industrial environment. Adding intelligent support and quizzes to the training can improve the training results.

8. Conclusions

This paper has investigated ARSG for industrial assembly from an operator perspective. Table 1 presents a summary of the findings.

<table>
<thead>
<tr>
<th>Category</th>
<th>Current status</th>
<th>Future challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly instructions</td>
<td>• Lack of standards</td>
<td>• Digitization</td>
</tr>
<tr>
<td></td>
<td>• Worker interest in individual and dynamic instructions</td>
<td>• Standardization</td>
</tr>
<tr>
<td>Human factors</td>
<td>• Video-based ARSG can cause efficiency losses</td>
<td>• Deeper evaluation of COM on ARSG</td>
</tr>
<tr>
<td></td>
<td>• Limited FOV in current ARSG</td>
<td>• Longitudinal tests of ARSG</td>
</tr>
<tr>
<td></td>
<td>• Interface potential safety risk</td>
<td>• Expansion of FOV</td>
</tr>
<tr>
<td></td>
<td>• Weight of ARSG should be kept a minimum</td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td>• Guidelines exists, focuses on presenting context-relevant information and limit disturbance of reality</td>
<td>• Sensors and visual recognition needs further improvements</td>
</tr>
<tr>
<td></td>
<td>• Sensors and visual recognition allows ARSG to interact with real world objects</td>
<td>• More verification and iteration of guidelines</td>
</tr>
</tbody>
</table>

It shows that there is currently a lack of standards in design of assembly instructions. Operators have also expressed interest in more customized and dynamic instructions as well as in using ARSG, and the increased complexity and updates leads to a need for dynamic instructions. The main future challenges regarding assembly instructions lie in improving standardization and digitization to enable ARSG compatibility.

In the human factors category it was found that video-based ARSG can cause efficiency losses and that there is a general limit of ARSG FOV. There are potential safety risks and the weight should be kept at a minimum, but the placement of the weight is also important. Future challenges identified are that weight and COM should be further evaluated and improved on, that more longitudinal user tests with ARSG are needed and that FOV in general needs to be expanded.

The current status in the design category is that available design guidelines focus on presenting context-relevant information and to limit disturbance of reality. Improvements in sensors and visual recognition has opened up more design alternatives by making it possible for ARSG to interact with real world objects. But future challenges lie in improving sensors and visual recognition. Current guidelines also needs to be further improved on and adapted to industrial settings.

In the support category the current status is that operators face complex and often changing tasks and that task routine is increasingly difficult to achieve. Future challenges lie in that this complexity is likely to increase in the future.

The current status in the training category is that many tests are simplified and not performed by operators. Few of the AR-studies has been done with ARSG and there has been few quality and training transfer measurements. Future challenges lie in performing more studies, mainly longitudinal ones.

The main contribution of this paper lies in that it gives a synthesized overview of what has been achieved and what still needs to be achieved when it comes to ARSG for operators within previously identified relevant categories. This overview will help to give an overall understanding of the current potential of ARSG as well as guide future improvements of ARSG for the use of industrial operators.

Future works include considering other relevant perspectives such as manufacturing engineering and technological maturity, further described in [6]. A more exhaustive review of the categories explored in this paper could also be beneficial, particularly validation which was only indirectly explored in this paper through the support and training categories.
References


[33] B. Schwerdtfeger, R. Reif, W. A. Günther and G. Klincker, Pick-by-vision: there is something to pick at the end of the augmented tunnel, Virtual reality, 2011, 15, pp. 213-223.


