Visual Assembling Guidance Using Augmented Reality

Anna Syberfeldt\textsuperscript{1}, Oscar Danielsson\textsuperscript{1}, Magnus Holm\textsuperscript{1} and Lihui Wang\textsuperscript{1,2}

\textsuperscript{1}University of Skövde, Skövde, Sweden.
\textsuperscript{2}Royal Institute of Technology, Stockholm, Sweden.
anna.syberfeldt@his.se, oscar.danielsson@his.se, magnus.holm@his.se, lihui.wang@iip.kth.se

Abstract

This paper describes a study of using the concept of augmented reality for supporting assembly line workers in carrying out their task optimally. By overlaying virtual information on real world objects – and thereby enhance the human’s perception of reality – augmented reality makes it possible to improve the visual guidance to the workers. In the study, a prototype system is developed based on the Oculus Rift platform and evaluated using a simulated assembling task. The main aim is to investigate user acceptance and how this can possible be improved.

Keywords: visual guidance, augmented reality, assembling, oculus rift

1 Introduction

Within industry, assembly line operators commonly face complex products structures and complex assembling sequences. The complexity, in combination with the pressure of completing the assembling within a minimum time frame but with maximum quality, makes it hard for the operators to act optimally. To aid the operators and support them to carry out the assembling task in the most efficient way, enhanced visual guidance can be used. This paper describes a study of using the concept of augmented reality (AR) for this purpose. With AR, artificial information about the environment and its objects can be overlaid on the real world in order to enhance the operator’s perception of reality. Formally, an AR system is defined as having the following features (Krevelen and Poelman, 2010):

(a) ability to combine real and virtual objects in a real environment,
(b) ability to register (align) real and virtual objects with each other, and
(c) ability to run interactively, in three dimensions, and in real time.

The technology enabling AR has advanced rapidly over the last years and there exist a number of real-world applications today, mainly within areas such as gaming, sports and tourism. Within the context of shop-floors, AR has been discussed for over 20 years, but there exist few practical
demonstrators. Tiefenbacher et al. (2014) discuss that AR has so far only partly succeed when it comes to industrial applications, mainly because the industrial setting is highly challenging and complex. We believe that a key factor for the success of AR on the industrial shop floor is acceptance by the users, and in this study we focus on this aspect with the aim of identifying factors that has potential to increase acceptance. With increased acceptance, we believe that AR is one step closer to become part of everyday assembling at the industrial shop-floors. The following aspects are central of our study, which altogether distinguish it from previous studies:

- **Focus on acceptability rather than performance**
  A literature review reveals that the majority of the existing studies within the field focus on the improved effectiveness gained with AR (i.e. being quantitative), which is also noticed by for example Dünser et al. (2008).

- **Focus on non-experienced users**
  Our study focus on junior operators that have no previous experience from the specific assembling task they are to carry out. This focus is motivated by the fact that unexperienced operators are those who can benefit most from AR from a training perspective. To a great extent, assembling is a repetitive task that is memorized over time and an experienced operator might therefore benefit less by AR support. Also previous studies indicate that AR is most efficient when the assembly task is difficult for the operator (see for example Wiedenmaier et. al, 2003), which most assembly tasks normally are for new operators.

- **Fully replicable assembling task that can be reproduced for benchmarking purposes**
  The assembling task used in the study is a 3D puzzle that this fully replicable and possible to use for future benchmarking and evaluations of AR solutions.

- **Prototype based on cheap, off-the-shelf consumer hardware**
  In the study, we develop an AR prototype that is based on the Oculus Rift platform which is a cheap consumer product that is easy to work with. As far as we are aware of, this is the first study that uses Oculus Rift for industrial assembling.

The study is part of the research carried out at the University of Skövde in Sweden, which aims to technically improve the industrial shop-floor and provide industrial operators with better tools in order to support the operators in making the right decisions and work optimally. The authors are certain that AR is a key to fulfill this aim, and also that the technique will be a part of all modern, high-tech shop floors of the future.

The next section continues by describing the approach used in the study for implementing AR. In Section 3, the equipment developed in the study is presented, followed by a description of the experiment performed in Section 4. In Section 5, results from the experiment are discussed. Section 6, finally, outlines conclusions from the study and possible future work.

## 2 Related work

AR for assembly support is a research community that was established already over 20 years ago. A summary of some of the most relevant studies within the community follows below in chronological order.

Caudell and Mizell (1992) investigated the use of a heads-up display in the assembling of aircraft wire bundles and developed an early prototype of AR. Curtis et al. (1999) later on investigated the prototype developed by Caudell and Mizell (1992) and performed additional user tests which indicated practical problems as well as acceptence problem. Reiners et al. (1999) developed an AR prototype
system for assembling door locks on cars which used an optical see-through interface for presenting instructions. Baird and Barfield (1999) evaluated the effectiveness of AR when assembling computer motherboards and showed that the users completed the assembling in shorter time when assisted with AR instructions compared to using ordinary displays or paper manuals. Boud et al. (1999) investigated the use of AR through a head-mounted see-through display for water pump assembly and found out that the users performed much quicker with this system compared to when only supported by two-dimensional engineering drawings. Tang et al. (2003) evaluated the effectiveness of using AR for assembling Duplo blocks and found out that the error rate was reduced by 82% when using the head mounted AR display compared to when the instructions was presented on an ordinary computer screen or printed out on paper. Zauner et al. (2003) developed a prototype system for AR in the assembling of furniture in which step-by-step instructions was given to the users who successfully managed to complete the assembling. Wiedenmaier et al. (2003) compared AR with paper instructions and expert guidance for a typical industrial assembling task and discovered that the assembling was completed in shortest time when the user was guided by an expert, followed by the use of AR support and in last place paper instructions. Nilsson and Johansson (2007) studied the use of AR as support in the assembling of a common medical device in a hospital and their analysis showed that the users were positive towards AR as a tool for instructions. Salonen and Sääski (2008) designed an AR system for the assembling of a 3D puzzle as well as a simulated industrial problem of assembling a tractor’s accessory’s power unit which used CAD models as graphical objects, but no user evaluations was performed as part of the study. Robertson et al. (2008) evaluated different AR solutions in the assembling of Lego blocks and found that a heads-up display with video see-through and context aware information was most efficient, but that the graphical information does not need to be located in the task area to be useful. Radkowski and Stritzke (2012) studied the use of AR for assembling by using virtual parts instead of real ones and they designed a solution based on a Microsoft Kinect video camera which was consider as intuitive to use according to the user tests. Hou et al. (2013) developed a prototype AR system for Lego assembling and found out that the system yielded shorter task completion times, less assembly errors, and lower total task compared to when using paper-based manuals. Radowski et al. (2015) investigated different visual features for AR-based assembly instructions and found out through user tests that the visual features used to explain a particular assembly operation must correspond to its relative difficulty level.

3 AR prototype

In this chapter, the design and implementation of the AR prototype system is described.

3.1 Selected approach

There exist a number of methods to implement AR. Krevelen and Poelman (2010) divide the various implementations into the categories presented in Table 1 below.

<table>
<thead>
<tr>
<th>Category</th>
<th>Divided into</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand-held</td>
<td>-</td>
</tr>
<tr>
<td>Head-worn</td>
<td>Retina projection, optical, video and projective</td>
</tr>
<tr>
<td>Spatial</td>
<td>Video, optical and projective</td>
</tr>
</tbody>
</table>

In this study, a head-worn implementation is selected. This is since such implementation frees the user's hands, which is seen as a necessary feature when considering shop-floors. It could be noted that
a spatial implementation also frees the user’s hands, but as this implementation requires fixed equipment in the working environment (Krevelen and Poelman, 2010), it is considered as too inflexible.

For realizing the head-worn implementation in the study, the choice stood between a video-based solution and an optical solution (retina projection and projective solutions were not considered for practical reasons). An optical see-through solution has the advantage of leaving the view of the real-world almost intact, attempting to merge a virtual image into the view of the real-world. A disadvantage of the solution, however, is that image updating at head moves often lags behind which affect the user experience negatively. Such visual lag is caused by, for example, communication or rendering delays and means that the virtual objects do not stay in the correct real-world position when the user moves. A video see-through solution has no such problem as the real view and the virtual view is merged into the same view, but on the other hand it has the disadvantage of blocking out the real world. In selecting between the two solutions, we consider it important to avoid the problem of latency in image updating that comes with the optical solution as this obviously is an annoying factor that will stand in the way when investigating user acceptance. Even normal head movements require an extremely fast and frequent image updating in order to avoid visual lag, which is hard to ensure even with the newest technology. To avoid the problem of lag, a video-based solution is chosen in the study. With this solution, the real world and the virtual world are merged into the same view, and the user's view is completely digitalized. In this way, the real world and the virtual world can be easily synchronized. One should be aware, however, that a video-based solution is not fully optimal in an industrial setting from a security perspective as the real world is blocked out and the user’s sight relies completely on hardware. However, while waiting for the head-worn AR solutions to become further developed and available at the market (for example, Google glasses* and C Wear†), the video-based approach is perfectly fine to use for experimenting in safe environments for research purposes.

In the next section, the hardware equipment used to implement the video-based solution is described.

3.2 Hardware equipment

The hardware platform used in the study is the first version of Oculus Rift (http://www.oculusvr.com/). Since Oculus Rift is developed for virtual reality where the real-world is totally blocked-out, modifications are needed for using it for AR. Such modifications have previously been made by Steptoe (2013). Steptoe mounted two web cameras in front of the Oculus Rift and showed their video streams on the Oculus's screens. With such solution, a video-based, digital view of the real world is achieved on which it is possible to place virtual objects. The same solution as proposed by Steptoe is used also in this study.

The modified Oculus Rift used in the study is shown in Figure 1. As can be seen in the picture, there are two sets of a bowed aluminum plates that can be slipped on from the side. These make it possible to adjust the cameras laterally which is important in order to compensate for different users having different distances between their eyes. Each aluminum plate has three holes in it and each hole has a nut screw. The screws are used to hold a plastic disc on which the cameras are glued. The plastic discs can be adjusted laterally by tightening or loosening the screws. For each screw hole in the plastic disc, there is a 25 mm slits that allows adjusting the cameras in height. The possibility to adjust the cameras in height has been considered as important in a previous study by Park et al. (2008). Park et al. also studied the effect of eye-hand coordination with different camera positions. They tested a number of different positions and found that a height displacement of 35 mm above or below the eyes gave a better user experience compared to when the cameras were placed at eye level.

* www.google.com/glass/
† http://www.penny.se/
3.3 Software

The software platform chosen for the project was Unity (www.unity3d.com) as this platform is easy to work with and fully supported by Oculus Rift. For implementing the AR, the software library Metaio (www.metaio.com) was chosen and used in combination with the software library Oculus Virtual Reality. Oculus Virtual Reality has built-in support for two cameras that renders digital content created in Unity, and to create the video connection all that had to be done was to catch data from the two web cameras and transform it into textures. Connecting Unity with Metaio was, however, a bit more challenging. Like Unity, Metaio uses two cameras but with a different purpose. In Metaio, one camera reads what is seen in the external camera (in our case, the web camera) and projects it on a plane. The second camera is used for the virtual projection. Based on what the first camera sees, the library is looking for patterns that are used for recognition. When a certain pattern is found, the position and rotation from the first camera’s perspective is determined. Based on this, the virtual objects are related to the pattern and rendered as a layer above the camera plane. The two layers, the external camera view and the virtual overlay, are then merged and rendered to the Oculus’ screens.

An interesting note to make is that there is an inevitable, small offset between what the user sees and real world due to delays causes by hardware communication and software processing. This cause certain a mismatch between the actual and perceived location of the user’s hand, but we have noticed that the brain very quickly adapt to the offset. The next section describes how the modified Oculus Rift has been used to in an experiment in order to evaluate the concept of AR in assembling.

4 User study

This chapter presents the user evaluation performed in the study and includes a description of the assembling task, the subject group, the experiment set-up and the evaluation.

4.1 Assembling task

The assembling task used in the study consists of assembling a three-dimensional puzzle with nine pieces (see Figure 2). The pieces are to be placed in a certain order and at specific positions, just like in industrial assembling. The task is relatively simple compared to real-world industrial assembling, but since the study does not mainly aim to assess the effectiveness gained with AR, but rather to investigate acceptance of the technology, we believe that the task is good enough to serve the purpose of the study. Similar assembling tasks for studying AR have also previously been successfully used by for example Sääski et al. (2008), Robertson et al. (2008) and Woodward et al. (2012), although these studies have a bit different focus.
When undertaking the assembling of the puzzle, the user wears the modified Oculus Rift. At the Oculus's screens, the piece to select next is highlighted in green and the place to put it is marked with the same shape and color. Figure 3 presents a screenshot illustrating the effect. In the screenshot, the user holds piece number one, which is overlaid with green color and pointed on with a virtual arrow. The piece is to be placed at the position marked with the same shape, color and number as the piece itself. As the reference for determining positions of objects a reference image is used. In this case, the reference image is a photo of a team of budo girls as can be seen in the screenshot. Several different random photos were tested and this particular one was the one most easily recognized by the computer program. A full demonstration of the AR function and a complete assembling of the puzzle can be found at http://youtu.be/tuP28sZ6EzM.

4.2 Subjects

Twelve participants were selected for the study, six of them from the student body of the university and the other six from the staff body. The age of the participants were ranging from 22 to 52 years. All participants were from an engineering background, but none of them had any previous experience of industrial manufacturing or assembling. Furthermore, none of the subjects had any previous experience of AR. Choosing subjects without any previous experience of assembling and AR is in line with the focus of the study which is unexperienced operators as those are the ones who can potentially benefit most from AR from a training perspective.
4.3 Experiment setup

In the experiment, half of the subjects were using the AR system and the other half was using only a traditional paper instruction (the two sub groups were randomly selected). An excerpt from the paper instruction is shown in Figure 4. The point-of-view and the information presented are as similar as possible in the AR system and the paper instructions.

![Figure 4](image)

Figure 4. Excerpt from paper instruction (the green boarders of a block indicates focus).

Before starting the assembling of the puzzle, all subjects were given a short introduction to the task. The six subjects using the Oculus Rift received basic information about the hardware and its functionality, but were not given any possibility to train on using the AR system (this is since the focus of the study is unexperienced users). During the assembling, no guidance was given to the subjects except when asked for it (which happened twice). All of the subjects were recorded on video while performing the assembling for enabling a detailed study of their behavior afterwards.

4.4 Evaluation

Dünser et al. (2008) describe that a common quantitative method used in AR evaluations is questionnaires, which is also the method used in this study to access the experience of the users. The questioner was filled in by the subjects after the assembling was completed. The questionnaire was developed based on the questionnaire used in Looser et. al (2007) and aims at assessing acceptability, including ease-of-use, satisfaction level, and approval. The questionnaire consists of the following seven questions, each graded on a Likert scale from 1 to 7 (1 = totally disagree, 7 = totally agree):

*Question 1:* I found the AR system easy to understand.
*Question 2:* I found it easy to use the AR system to place the pieces.
*Question 3:* I felt that I performed quickly with the AR system.
*Question 4:* If I had to use an AR system like this on a regular basis, this is a technique I would appreciate having available.
*Question 5:* I found the AR system physically demanding.
*Question 6:* I found the AR system mentally demanding.
*Question 7:* I found the AR system frustrating.

Besides the questionnaire, objective measurements of task completion time and error rate were also performed. Both completing time and error rate was measured manually. Evaluating the effectiveness gained with the AR system is not of main focus of the study, indeed, but we believe that objective measurements has the potential to add additional insights it might and be relevant for future benchmarking. It should, however, be pointed out that proving the efficiency of AR as such is not a goal of the study; this have already been done in many previous studies (see Chapter 2) using assembling tasks that are more suitable (complex) for that specific purpose.
5 Results

This chapter presents the results of the evaluation of the AR system.

5.1 Subjective evaluation

Figure 6 presented the average results from the questions in the questionnaire. As can be seen in the figure, the AR system was considered easy to understand – almost as easy as the paper instructions (question 1). We believe that the system being easy to understand is an important factor for acceptability, since it is usually hard to motivate new users to try out a system if it is perceived as hard to understand. However, when it comes to ease-of-use (question 2), the paper instructions were considered significantly easier to use. Also when it comes to perceived efficiency the paper instructions were superior. From an acceptability perspective, this can be problematic since it might be hard to convince the users to continue using the system if it is considered as hard or inefficient to use. One should, however, keep in mind that the subjects were completely new to the technology and they might have considered the system easier to use and more efficient if given some more training. In relation to the answers on question 1-3, it is interesting to notice that the subjects to a high degree would appreciate having access to an AR system on a daily basis (question 4). This we believe is a strong indication of high acceptance.

When it comes to question 5, the subjects experienced the AR system as physically demanding. The main cause mentioned for the negative experience was the heavily deteriorated vision that followed by using the equipment. In fact, this is no surprise since the resolution of the Oculus Rift’s screens is only 640 x 800 pixels. The user cannot only see the individual pixels, but the pixels are also placed with a small gap between them. The result is a quite blurry sight with a black grid net laid upon everything the user sees, which is disturbing. Furthermore, the sight angle is significantly reduced as the angle of the web camera (which is used record the real-world) is limited. However, it is important to keep in mind that the Oculus Rift at the moment is at a very early stage and that the hardware will have considerable better screen resolution and refresh rate when released as a commercial product. The next version of the Oculus Rift has recently been released with a screen resolution of 960 x 1080 pixels, and the commercial version coming later on will have even higher resolution. As the technology advances, the user experience will increase significantly.

The perceived mental strain and frustration (questions 6 and 7) were approximately the same with the AR system and the paper instructions.

![Figure 6. Questions and average results (1 = totally disagree, 7 = totally agree).](image)
5.2 Objective evaluation

In this subchapter, the measured error frequency and time consumption are presented. Due to the low sample size, no statistical analysis has been performed.

4.1.1 Error frequency

The participants using the AR system all succeeded in completing the assembling task without any errors in the final placement of a piece. This is expected since the system is designed to force the user to do the right thing, and the user is not allowed to proceed with the next piece until the first is placed correctly. In comparison, two of the six subjects using paper instructions placed pieces incorrectly, which were detected in later steps of the assembling process and then corrected.

4.1.2 Time consumption

The amount of time needed for carrying out the assembling varied significantly between the subjects using the paper instructions and the ones using the AR system. With the paper-based instructions, the fastest assembling was performed within 1 minute and 1 second, and the slowest within 1 minute and 28 seconds. With the AR system, the fastest assembling was performed within 3 minutes and 37 seconds, while the slowest assembling took about three times as long - it was completed within 10 minutes and 30 seconds. The significant difference in amount of time needed for carrying out the task indicates a spread among the subjects regarding their ability to adapt to the equipment. A familiarity with the equipment being advantageous becomes clear when studying the time consumption of individual pieces. As shown in Figure 5, the first two pieces took considerable longer time to place correctly compared to the rest.

![Figure 5](image)

Figure 5. Time consumption of each of the nine pieces of the puzzle (average values).

After placing the first two pieces of the puzzle, the subjects using the AR system had learned the basics of how to use the equipment and were able to perform faster. The single most important aspect for the subjects to learn in order to perform the task efficiently was to retain the reference image (the photo of the budo girls) in the camera's field of view. When the reference image gets out of sight, no virtual information can be shown and the user then gets completely lost. All of the subjects experienced this couple of times, especially with the first two pieces before learning how to avoid it.
6 Conclusions and future work

Although the concept of AR has been discussed for decades, it has so far not yet made its breakthrough at the industrial shop floor. We believe that a key factor for such breakthrough is high acceptance by the users, and if this can be reached we believe that AR is one step closer to become part of everyday assembling at the industrial shop-floors. In this study, we have looked further into acceptability with a focus on junior operators that have no previous experience from the specific assembling task they are to carry out. An AR prototype based on the Oculus Rift platform was developed as part of the study and used for assembling a 3D puzzle. The assembling task is fully replicable and possible for others to use for benchmarking.

An analysis of the results from the user study has given us a number of insights in how acceptability of AR systems can possible be improved. Some of the key factors we believe are the following:

- **Identify assembling tasks of high enough complexity**
  The complexity of the assembling task must be high enough for the user to feel that it is worth using the AR system. Preferably, the task should be so hard that it is almost impossible to complete without the AR system, at least for a junior operator.

- **Ensure improved efficiency**
  It must be proved that the AR system makes the user more efficient, otherwise the user will see no meaning in using it. Proper training of new users is also important in order to make sure that the system is used in the most efficient way.

- **Aim for a perfect system**
  A system that comes without flaws is significantly easier to accept than a system that contains apparent imperfections. Building a perfect system in general is no easy task, and it becomes even harder when it comes to AR as the enabling technologies are not yet fully mature.

- **Emphasize the advantages**
  When introducing AR to new users, make sure to fully “sell” the system and emphasize all its advantages the user’s perspective. One important advantage to rise is the built-in control against assembling errors that comes with the AR system. This type of built-in control helps to protect against human mistakes and is very valuable in industry where maximum quality is sought for.

In the near future, the authors intend to extend the study with the new Oculus Rift as well as with other hardware solutions for AR. Planned future work also includes involving a larger number of participants in the experiment and to introduce several other assembling tasks with higher complexity. Preferably, some of these assembling tasks should include steps that require the use of tools as this introduce additional challenges when it comes to AR support. It would also be advantageous to study ergonomic aspects as part of future studies, for example related to head and hand movements. Body coordination with and without equipment would also be interesting to investigate.
References


