Final Report – Group 6

Names: Sreeya Basu Roy, Mohaddeseh Basiri, Eric Samikwa, Yichen Cao

Title of project: Instruction Aider (AR Instructor)

Supervisor: Mario Romero Vega

1. Abstract:

One of the most lucrative applications of Augmented Reality (AR) technology is in the traditional manual assembly in manufacturing and home using to assemble devices. In order to minimize errors of assembling process, we present an augmented reality-based assistance application which helps users how to assemble an electronic circuit. In this way, we aim to reduce the overall effort and expertise required and time taken in assembling the circuit by using multimodal interfaces and interactions. This application is compared with traditional manual assembly (i.e. using textual description) regarding to the time required for assembly and user acceptance. The evaluation results indicate that assistance application has better performance by 19.65%. Furthermore, a majority of the users preferred using the application over reading an instruction manual.

Keywords:

Augmented reality, Assembly process, Multimodal interactions.

2. Introduction:

Human beings have experienced an unseen growth in technology within the recent years, which is expected to be dramatically increased by the next industry revolution, i.e. Industry 4.0 [1]. For each industrial revolution, the main driver has been increasing the productivity [2]. It is expected that the Industry 4.0 will not only significantly improve the productivity by introducing revolutionary solutions like smart and autonomous factories, but it will also significantly improve the design, engineering, training, and maintenance processes by leveraging everything connectivity and learning from the mass volume of available data [3]. Among these diverse aspects, let us focus on the training procedure. The rapid changes of design processes, software modules, and products mandate mass training of technicians and employees in different industries in order to help them acquire new skills to assemble and maintain new products. However, training employees on such a large scale is often time and money consuming mainly because assembly and maintenance tasks can be extremely detailed and complex. In some cases, experts might have to travel from distant places in order to conduct training workshops. Therefore, finding modern and efficient methods to train technicians in assembly and maintenance tasks is the need of the hours. Thanks to the significant changes in the information and communication technology field, virtual and augmented reality, denoted by VR and AR respectively, have attracted profound attention as solutions for addressing such challenges. To be more specific, the traditional suggestion and guidance provided by senior technicians to juniors could be unified and packed as a software application, hereafter also called Instruction Aider, to help juniors find their way in their job. The usefulness of such solution is not only limited to the training process, and it could be also used by individual consumers who buy certain products and need to assemble them at home, e.g. one may refer to the shopping model of IKEA for home furniture. The traditional user manuals that come with the product have textual descriptions which may be interpreted in different ways depending on the context of understanding of the user. Another issue that many consumers face is that the instruction manual is often written in another language than that understood by the consumer, which must then be translated by the user using online tools. Our project, aims to combat such day to day disadvantages. As the name aptly suggests, AR actually augments consumer experience by superimposing a virtual environment onto a real machine. This can be achieved by either using visual headsets or using handheld, portable devices (e.g. smartphones).

In this project, we have developed an augmented reality-based application on a smartphone which can be used for training by simply pointing the camera of the trainee's smartphone onto a specific part/port of the device to be joined to another part/port. The assistance application communicates with the trainee with the help of sonic and visual feedback and in this way the trainee will be aware of the connections to make.

The reminder of this report is organized as follows. In Section 3, the related works are reviewed. In Section 4, setup and methodology is described. Strengths and weaknesses of assistance application is provided in Sections 5 and 6. Evaluation scheme and future works are presented in section 7 and 8. Finally, concluding remarks is in section 9.

3. Related Works:

In recent years, AR has attracted profound attention in design, assembly, and maintenance of products, in both academia and industry. Here, we focus on applying AR for the assembling process in order to improve hands-on experiences in using assembly and maintenance manuals.

Boeing was one of the first companies that deployed an AR-based system to assemble different pieces of airplanes [4], [5]. They aimed to improve workforce efficiency and reduce the costs of using traditional templates and manuals in the manufacturing and assembling process. Zuaner et al. [6] implemented an AR application for assembling furniture leveraging the ARToolkit, which tracks and recognizes the components. Yuan et al. [7] developed a prototype AR-based assembly instructor system that features a virtual interactive panel. In this panel, an assembly operator can easily pursue the pre-defined assembly procedure without requiring any sensors or markers on the assembly components. Reiners et al. [8] represented an AR-assisted system that displays instructions for assembling a car door. This system leverages a tracked monocular optical see-through (OST) head-worn display (HWD) presenting instructions to the operator for assembling different components. Pathomaree and Charoenseang [9] implemented an AR-assisted assembly training system to improve the skill transfer in the assembly task. Their experimental results indicated that not only the AR-based training system had a high transferability level and high transfer success ratio, but also the assembly completion time and the number of assembly steps reduced in the proposed system.

Tang et al. [10] investigated the AR success in assembling toys blocks. Their results showed that in comparison with the traditional approach, users made fewer mistakes when following displayed instructions on the AR-powered display. Robertson et al. [11] found that displaying fully registered instructions on a tracked binocular video see-through (VST) HWD can trigger users to assemble Lego blocks more quickly than non-registered alternatives. Furthermore, Baired and Barfield [12] examined the effectiveness of augmented reality displays in assembling a computer motherboard in comparison with paper manual and computer-aided instructions. Their experimental results demonstrated that the augmented reality outperforms the paper instruction manual and the computer-aided instructions. Furthermore, the assembly process in see-through augmented reality display took the lowest time and opaque augmented reality display, computer-aided instruction, and the paper instructions were in the following ranks respectively. In addition, people made fewer errors in using augmented reality during the assembly process.

4. Setup and Methodology:

We developed the assistance application on the Android smartphone for helping users to assemble components on a PCB (In the demo, we used paper for replacing the PCB). When users are assembling components, they can open the application on their smartphone and use the camera to scan the mark on the board. Then, the 3D virtual models of components will be displayed in the positions where the real components should be on the board via the screen. Users can follow the 3D models to find the position of components without looking up in documents. Meanwhile, if users want to focus on one single component, they can speak and ask the application where the exact component is. Then the application will only display the selected component. All the components are attached with marks. When users put the real component on the correct position, the application can detect the mark and will play sounds as feedback for a successful assembling.

4.1. Environment:

The application runs on the Android system. The display of the assembly procedure is based on AR technology. Unity3D Engine with Vuforia SDK which provides a simple approach to implement AR is used for development. The 3D models used for displaying the PCB layout is built on TinkerCAD which is an online design platform and is very easy to learn and use. The speech recognition is implemented via Google application programming interface (API).

4.2. Development:

The process of development was divided into the following steps.

1) Modeling components: In the first step, we built the 3D models of the electrical components (Mainly capacitors, transistors, and IC chips) via ThinkerCAD. Since we do not need too many details and complex operations on the components, modeling on ThinkerCAD is enough.

2) Design of marks: We designed marks for the camera to recognize so that the application can know which board and which components the users are going to assemble. The marks were attached on both board and components. The marks of components should be different from each other, meanwhile, they should have enough complexity and features so that they can be well recognized. It is not possible to design all the marks one by one since there are many marks to design. So, we transformed the text information of components to QR codes and used the QR codes as the marks.

3) Recognizing and displaying: The process of recognition is based on the Vuforia SDK which provides an image target recognition approach. The images are our designed marks uploaded to the Vuforia developer database. The images are added to the application as a Vuforia image target. When the camera scans the mark, the application can automatically recognize the image and display the 3D models of components.

4) Speech instruction recognition: The basic function of speech of recognition is based on Google API. When the user asks where the component is, the application transforms the speech information to text information via Google API. When the application gets the text information, it searches keywords (including “where”, “position”, and names of components) in the text and figures out the component that the user needs. Then the application will hide other component models and leave the one that the user needs. When the user says “reset”, all the models would be displayed again.

5) Correctness detection: When the user is placing the component, the application would recognize the mark on the component, and compare the coordinate of the mark with the coordinate of the position that the component should be. When the component is placed on the correct position, the application would make some sound feedbacks for notifying that the component has been correctly placed.

4.3. Theoretical Analysis:

Using AR to assemble a circuit or equipment is considerably easier than using textual manuals since it is easier to map real components to 3D models other than having to interpret written text and 2D images on paper [13]. A user can easily understand or remember how two parts are to be connected after seeing a model of a component in 3D model placed on a real motherboard using AR. This shifts complexity of assembling from user to the system as the cognitive load of translating abstracted instructions onto reality is reduced [14]. Using speech as one of the modalities for interact with the system makes it feel more natural when using it [15]. Natural language capability used in speech recognition is necessary as it does not require a user to go through any training to use it [15], all the user is required to know is the names of the components. For example, a user can say “where is [component name]” and the system displays a 3D model of the component on the position where it supposed to be on the real board. Since the user is assembling the circuit, the assumption is that he/she may need to use two hands hence speech is more convenient. Sinification has been used to guide a user when placing components on the board. A sonic feedback given when a user places a component on the correct position feels more natural since users are already familiar with sound for “ok” or success notifications in their smart phones or computers [16].

The three modalities have been used as complementary modalities to achieve ease of use and less processing complexity for the user [17]. Using one modality such as AR only could not solve the problem efficiently. Users would be required to be constantly touching the phone for them to display a specific component, which can lead to increase in time required to assemble the circuit as one may be using one hand. AR only would also require a user to be constantly looking at the phone display when placing the components as there will be no sonic feedback to alert the user of a correct position. Using speech as the only modality can result in making the system very uncomfortable as it very difficult to describe the position or alignment of components by speech [18]. Hence the multi-modality aspect of the systems is crucial to its efficiency and usability.

5. Strengths:

By studying related works parts, one observes that the previous works has been focused on using a limited number of modalities in assisting operators, and the AR-assisted assembly has been mainly focused to the mechanical challenges. Backed to the state-of-the-art research on AR-assisted assembly, in this work we leverage augmented reality in assembling electronic components. Apart from reducing the time for assembling systems, the proposed solution also results in reducing the cost which is wasted otherwise in failure of electronic equipment. While the previous part has been focused on using a limited number of modalities in assisting operators, our work features using multimodalities; including speech for input, and sound, vision, and augmented reality for output. Furthermore, our work benefits from Google speech application programming interface (API) for recognizing the input commands, which enables users in different languages to communicate with the application. Finally, instead of implementing the AR-assisted solution over special hardware, e.g. OST HWD, our solution has been implemented in smartphone, which is widely available, and hence, can provide a cost-efficient option to users.

6. Weaknesses:

Some weaknesses of assistance application can be mentioned as following:

1. Marks:

   The marks used for Vuforia image target recognition must be big and clear enough. The shape of the marks we used is square with 1.5cm on each side so that the application can recognize it easily. But the marks become much bigger than the components. They totally cover the components and make it difficult to see the components. Besides, when the mark moves too fast, the application would lose the trace of the mark.

2. Speech recognition:

   We used Google speech-to-text API to implement speech recognition. It works well except when recognizing the name of the components. Some component names like “C1”, “TP” are difficult to recognize since they are not natural words.

3. Correctness detection:

   We planned to compute and trace the distance between the mark on the component and the correct position of the component, and give sound feedbacks when the mark is placed on the correct position. But it turns out that when the camera is set at different angles above the board, the computed distance would change a lot which influences the accuracy of detection.

7. Evaluation scheme:

To verify if the main aims of the project are met, the correctness of the developed application was evaluated. This was a non user-based evaluation of the application to determine whether all required functionalities are present and are performing as expected. The ability of the application to detect markers on a 3D object from different angles and in different light intensity was examined. The ability of the application to direct a user in assembling equipment using the three output modalities was examined. Since the main aim of the application is to reduce the time taken in assembling equipment, a user-based evaluation was also conducted. This involved conducting an experiment with real users who are not part of the developing team. To determine if the application reduces the complexity of instructions and the time taken in assembling equipment, a qualitative approach used. User experience and acceptance of the application was also investigated in the process. The analysis was based on observations and a set of questions which each user answered after using the application. Learnability of the application was also determined by observations.

7.1. Test and evaluation:

a) Function test: In the function test, three functions of the application were mainly tested.

Firstly, the ability of the application to recognize markers was tested. Second, the ability of the application to shows the virtual components on the precise and correct position. Thirdly, the speech recognition capability of the application was tested, thus if the application is able to locate position of a specific component based on a given speech command.

b) Usability evaluation: In the evaluation phase, we organized 10 people as users to test the application. The users were asked to assemble a circuit comprised of eight components on a circuit board. The users assembled the circuit with the help of our application and the time taken and feedback was collected. The users were then asked to assemble the circuit with only text instruction from a written manual and the outcome was also recorded. Table 1 shows the time taken to assemble the circuit using the two procedures. Figure 1 shows the graph of time taken to assemble the circuit using the two procedures.

Name of user	Time Taken Using the App	Time Taken Using the

	(sec)	Instruction Manual (sec)
Madhurya	200	300
Keerthana	192	308
Shreya	224	280
Shipra	250	280
Gautham	232	260
Shrinish	192	296
Abhishek	240	308
Shekhar	280	288
Govind	300	280
Sreeya	220	300

Table 1: Time taken to assemble the circuit using the two procedures.

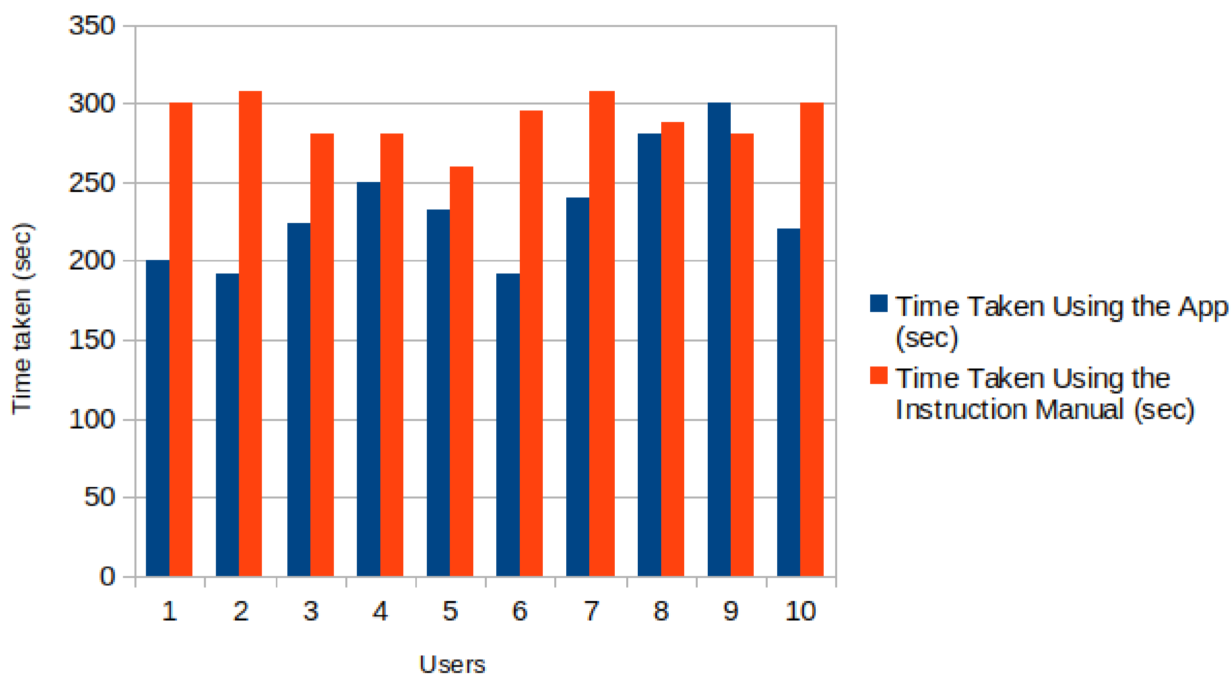

Figure 1: Graph of time taken to assemble the circuit using the two procedures.

The evaluation results indicate that assembling the circuit with the help of the instruction aider application took less time (19.65 percent) than assembling the circuit using written instruction manual. On average, assembling the circuit using the instruction aider application was about 60 seconds faster than the textual manual. Furthermore, the number of errors in selecting components and finding their locations reduced by deploying application and users had better experience in assembling the circuit by instruction aider application.

8. Future works:

To improve assistance application, following steps have been planned to carry out:

1. Investigate the combination of augmented reality with intelligent systems [20].
2. Redesign the marks:

   Try other types of marks which can be recognized in a smaller size so that the marks can be attached to the components without covering the whole components.
3. Improve speech recognition:

   We can provide a set of words that are likely to be spoken (for example, the names of components) to the speech API. It would significantly improve the accuracy of speech recognition especially the recognition of component names.
4. Improve correctness detection:

   For eliminating the influence of camera position, we can take into account the transformation of coordinates and angle of view to compute the position of the mark and the distance between the mark and correct position.

9. Conclusion:

In this project an application that can be used for training and assembling was created. Due to the projects time frame, the project focused on assembly of one type of equipment, an electric circuit in which the electronic components and the board were identified by markers. Furthermore, a literature review was conducted to determine how well different modalities can be combined to help make the system efficient and usable. The final system was evaluated by testing it with different users who were not part of the development team. From the results, it is clear that the instruction aider reduced the time taken about 19.65% to assemble the circuit. Furthermore, the users preferred using the application than reading the instructions from the manual due to its ease of use. This was achieved using the multimodal system which is made of AR, speech and sonification. It was also noted that using the modalities separately could not achieve better results. Due to other challenges which were faced during the design of the system, it was concluded that improvements can be done by redesigning markers that can work on small components, improving speech recognition to detect different accents easily and improving the detection of correct position in AR, in the future. Furthermore, investigating on combination of augmented reality and intelligent systems will be followed.

Link to the video:

https://vimeo.com/311064271

References:

[1]. Lasi, Heiner, Peter Fettke, Hans-Georg Kemper, Thomas Feld, and Michael Hoffmann. "Industry 4.0." Business & Information Systems Engineering 6, no. 4 (2014): 239-242.
[2]. Brettel, Malte, Niklas Friederichsen, Michael Keller, and Marius Rosenberg. "How virtualization, decentralization and network building change the manufacturing landscape: An Industry 4.0 Perspective." International Journal of Mechanical, Industrial Science and Engineering 8, no. 1 (2014): 37-44.
[3]. Lee, Jay, Hung-An Kao, and Shanhu Yang. "Service innovation and smart analytics for industry 4.0 and big data environment." Procedia Cirp 16 (2014): 3-8.
[4]. Caudell, Thomas P., and David W. Mizell. "Augmented reality: An application of heads-up display technology to manual manufacturing processes." In System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on, vol. 2, pp. 659-669. IEEE, 1992.
[5]. Curtis, Dan. "Several Devils in the Details-Making an AR Application Work in the Airplane Factory." In Proc. Int'l Workshop Augmented Reality, 1998. 1998.
[6]. Zauner, Jürgen, Michael Haller, Alexander Brandl, and Werner Hartmann. "Authoring of a mixed reality assembly instructor for hierarchical structures." In Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, p. 237. IEEE Computer Society, 2003.
[7]. Yuan, M. L., Soh-Khim Ong, and Andrew YC Nee. "The virtual interaction panel: an easy control tool in augmented reality systems." Computer Animation and Virtual Worlds 15, no. 3-4 (2004): 425-432.
[8]. Reiners, Dirk, Didier Stricker, Gudrun Klinker, and Stefan Müller. "Augmented reality for construction tasks: Doorlock assembly." Proc. IEEE and ACM IWAR 98, no. 1 (1998): 31-46.
[9]. Pathomaree, Nattapol, and Siam Charoenseang. "Augmented reality for skill transfer in assembly task." In Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on, pp. 500-504. IEEE, 2005.
[10]. Tang, Arthur, Charles Owen, Frank Biocca, and Weimin Mou. "Comparative effectiveness of augmented reality in object assembly." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 73-80. ACM, 2003.
[11]. Robertson, Cindy M., Blair MacIntyre, and Bruce N. Walker. "An evaluation of graphical context when the graphics are outside of the task area." In Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on, pp. 73-76. IEEE, 2008.
[12]. Baird, Kevin M., and Woodrow Barfield. "Evaluating the effectiveness of augmented reality displays for a manual assembly task." Virtual Reality 4, no. 4 (1999): 250-259.
[13]. Boud, A.C., Haniff, D.J, Baber, C. and Steiner, S.J., Virtual reality and augmented reality as a training tool for assembly tasks. Inter. Conf. on Info. and Visualization, (1999)
[14]. A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In CHI'03: Proc. Int'l Conf. on Human Factors in Computing Systems, pp. 73–80, Ft. Lauderdale, FL, USA, 2003. ACM Press. ISBN 1-58113-630-7.
[15]. T. Komatsu, A. Ustunomiya, K. Suzuki, K. Ueda, K. Hiraki and N. Oka (2005). Experiments Toward a Mutual Adaptive Speech Interface That Adopts the Cognitive Features

Humans Use for Communication and Induces and Exploits Users' Adaptations, International Journal of Human–Computer Interaction, 18:3, 243-268, DOI: 10.1207/s15327590ijhc1803_1
[16]. T. Hermann and A. Hunt. An introduction to interactive sonification. IEEE Multimedia, 12(2):20–24,2005. Editorial, special issue on interactive sonification
[17]. Bouchet, J., Nigay, L. ICARE: A Component-Based Approach for the Design and Development of Multimodal Interfaces. In Extended Ab- stracts CHI'04 (2004). ACM Press, 1325-1328
[18]. Bernd Schwald, Blandine de Laval: An Augmented Reality System for Training and Assistance to Maintenance in the Industrial Context, (2003)
[19]. De Crescenzio, F., Fantini, M., Persiani, F., Di Stefano, L., Azzari, P., & Salti, S: Augmented reality for aircraft maintenance training and operations support. IEEE Computer Graphics and Applications, 31, 96–101 9, (2011)
[20]. Westerfield, Giles, Antonija Mitrovic, and Mark Billinghurst. "Intelligent augmented reality training for motherboard assembly." International Journal of Artificial Intelligence in Education 25, no. 1 (2015): 157-172.