Comparing Spatial and Mobile Augmented Reality for Guiding Assembling Procedures with Task Validation

João Alves[1], Bernardo Marques[1], Miguel Oliveira[2,3], Tiago Araújo[1,4], Paulo Dias[1], Beatriz Sousa Santos[1]

[1] *DETI, IEETA*
University of Aveiro
Aveiro, Portugal

[2] *DEM, IEETA, TEMA*
University of Aveiro
Aveiro, Portugal

[3] *INESC-TEC*
Porto, Portugal

[4] *PPGCC*
Federal University of Pará
Belém, Brasil

***Abstract*—Assembly tasks are a common situation in many industrial applications. These tasks are often presented on paper or digital manuals containing instructions, photos or diagrams to guide an assembly sequence. While some Augmented Reality (AR) systems have also been proposed to support these processes, only a few track the state of the assembling procedure, validating the process in real-time. In this work, we propose two different AR-based (mobile and spatial AR) methods with real-time validation to provide assistance to users during the execution of an assembly process. The validation process uses computer vision techniques to keep track of the state of the assembly sequence, verifying the completion of each stage and providing information at the end of the assembly. A controlled experiment was used to compare the performance, ease of use, and acceptance of the two AR-based methods proposed. Participants were significantly faster and made fewer errors using the Spatial AR condition. Besides, participants also preferred this condition. In addition, Nasa TLX rating showed that the Spatial AR condition had a slightly lower cognitive load on the participants.**

Keywords — Mobile Augmented Reality, Spatial Augmented Reality, Assembly Guidance, Computer vision

I. INTRODUCTION

While some assembly processes are automated, a significant number of assembly operations still require manual intervention due to their complexity. Conventional assembly processes often resort to descriptive instructions on paper or in digital format (photos or diagrams) to guide the assembly sequence. For this reason, users are required to map between these instructions and the actions to be performed on real objects, without getting any feedback or help [1].

Through Augmented Reality (AR) it is possible to display digital contextual information [2] in the field of view of the subject performing an assembly process, such as step-by-step instructions, 3D illustrations, or other relevant data [3]. Some AR-based systems have been proposed to assist in assembly guidance. Still, only a reduced number is able to detect errors in real-time and track the current state of the assembly sequence [1][4].

This paper presents a prototype based on two AR-based methods (mobile AR and Spatial AR) that implements validation using computer vision techniques. One objective is to close the AR loop, providing assistance to users during assembly and giving them feedback about their actions. To compare the methods, we performed a study to evaluate performance, ease of use and acceptance of the two methods and thus evaluate the potential of two different AR configurations with validation in assembly scenarios.

The paper briefly presents related works, describes the experimental setup and design, presents and discusses the results obtained, and draws conclusions and ideas for future work.

II. AUGMENTED REALITY IN ASSEMBLY GUIDANCE

Several research works have been focusing on AR-based systems using different methods to assist assembly procedures. Tang et al. (2003) explored the use of AR presented using a Head Mounted Display (HMD) and performed an experiment to evaluate the performance of AR instructions in an assembly task. A user evaluation with 75 participants using LEGO Duplo blocks was conducted. Participants were required to complete an assembly task following the instructions presented in different media: 1- Printed media, 2- Instructions on a monitor, 3- Instructions on a see-through HMD, 4- Spatially registered AR instructions on a see-through HMD. The study provided evidence that AR-based systems can improve task performance (lowest average time and number of errors) and reduce mental workload on assembly tasks compared to other media [5].

Khuong et al. (2014) presented an AR-based system using the combination of a Head Mounted Display (HMD) with a Kinect sensor to display guidance and error detection information corresponding to the assembly status. A preliminary study was conducted with 24 participants using LEGO building blocks. Results showed participants preferred

To all participants involved in the case studies, thanks for collaborating in providing essential data and relevant feedback. The present study was developed in the scope of Produtech – SIF – Solutions for the Industry of the Future [POCI-01- 0247-FEDER-024541]. This study was also supported by IEETA - Institute of Electronics and Informatics Engineering of Aveiro, funded by National Funds through the FCT - Foundation for Science and Technology, in the context of the project UID/CEC/00127/2013.

978-1-5386-5346-6/18/$31.00 ©2019 IEEE

the visualization mode displaying information next to real objects in comparison to the traditional direct overlay [3].

Nishihara and Okamoto (2015) presented one of the first AR-based system using a mobile device (instead of the traditional HMD), combined with object recognition through image processing for assembly guidance. The system focused on object recognition and was tested using a Pentomino Puzzle. It obtained a 98% success rate using a recognition based on the combination of the Canny Edge algorithm, morphology closing and border following algorithm. No formal evaluation with users was presented. Notwithstanding, the authors claim that AR-based research works can be structured according to 4 main aspects, including type of display, type of tracking system, user interaction and platform [6].

Loch et al. (2016) evaluated an AR-based assistance system integrated in a manual workstation using a camera to track the user's workflow and follow the assembly tasks stage automatically. A study was conducted with 17 students using LEGO building blocks. The system was compared with video-based assistance regarding performance, user acceptance and mental workload. Results showed improvements in accuracy, task performance and reduction in the number of errors and task time when AR was used [7].

Hence, literature acknowledges AR-based technologies can be interesting tools to minimize errors and decrease user's mental workload in comparison to other methods (e.g. computer assisted instructions, paper manuals) [5][7]. These advantages, combined with the emergence of more affordable and powerful devices, such as mobile devices and AR head-sets [8], make AR especially suitable for assembly guidance [9][10][11]. Although a number of prototype systems using AR have been proposed, few monitor the assembly status and detect errors in real-time; yet, this might be a useful feature in such systems and worth integration in the paradigm. Other relevant aspects are the type of platform and display. These issues motivated us to integrate a validation mechanism into different AR-based approaches to assembly guidance and compare them to better understand benefits and limitations.

III. Prototype

To compare several AR-based methods we implemented a prototype to guide users through an assembly task displaying the virtual representation of the real objects. To show the assembly steps to the user we used two devices: a mobile device and a projector. Figure 1 shows the components of the prototype which are a computer, a mobile device, an RGB-D camera, a projector, a local storage and an AR marker.

The platform architecture is based on a Server-Client approach, focusing on simplicity and scalability. The architecture exchange messages (based on specific events) between applications to trigger updates on the interfaces of the devices, as long as they are connected to the same wireless network. The server connected to the RGB-D camera is responsible for capturing the workspace area where the user is assembling the structure validating if the building blocks are correctly placed. The server also contains a local storage, with the 3D and 2D templates used in the validation process. The client is responsible for showing the augmented world with the next building block in its target pose, using a mobile device or a projector.

We used Unity, Vuforia and ROS to develop the system. Unity and Vuforia were used to develop the AR application and define the pose of the pieces to be used in the assembly. To monitor the position of the blocks, we used ROS and its libraries. We used the ROS#1 plugin as well to communicate between these two platforms (Unity and ROS).

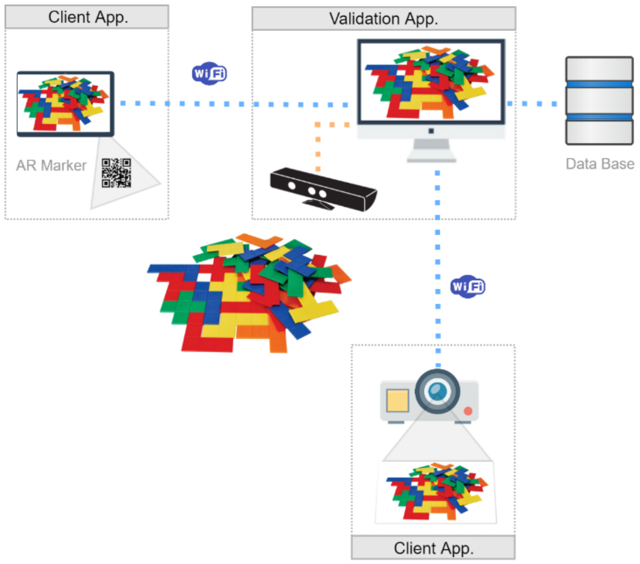

Figure 1 – Prototype setup with its modules and relationships.

A. AR-based Display Methods

As mentioned before, we propose two AR methods to assist assembly procedures with validation. One of them uses a mobile device and the other one uses a projector. The perspective in which the assembly instructions are shown are different for each method. To compare the methods, we developed an application to assist in building a puzzle with pieces of different shapes, placing them in predefined poses to construct the final assembly. The prototype, using an RGB-D camera, detects the pose of the pieces on top of the table and checks if the user has placed them correctly.

1) Method 1 – Mobile Augmented Reality

The first method uses a mobile device to augment the environment with the next piece to be assembled in its target pose, as depicted in Figure 2. The validation of the correct position of the piece is used to trigger the transition between assembly stages. This method needs a previous calibration step to provide the transformation between the RGB-D camera used for validation and the marker used to determine the pose of the mobile device. This implies an extrinsic calibration between the RGB-D camera and the assembly parts in their respective predefined poses and a calibration between a natural marker and the assembly building blocks.

[1] https://github.com/siemens/ros-sharp

Figure 2 – Method 1 – Mobile Augmented Reality with validation.

2) Method 2 – Spatial Augmented Reality

The second method is similar to the first one, but instead of a mobile device, a projector is used to display the assembly instructions as virtual pieces. This particular setup as well as its projection are shown in Figure 3. Unlike the first method, the projection needs a flat surface to present the information, preventing us from having a tridimensional structure.

This method also needs a previous offline phase to perform a calibration between the projector and the RGB-D camera. The calibration process uses four projected pre-defined markers, with a known size, to determine the position of the working area and its dimensions in the world. Knowing these point coordinates and size of the pieces, it is possible to determine the grid dimensions and consequently the size of the virtual pieces.

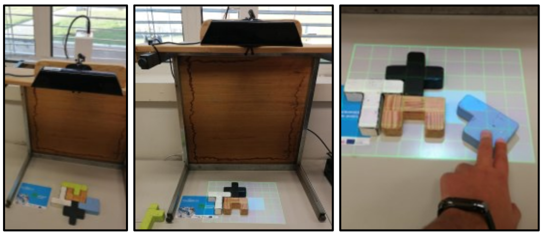

Figure 3 – Method 2 – Spatial Augmented Reality with Validation. Setup with the overlaid grid and the next virtual piece to place (left) and detail (right).

B. *Validation through Object Recognition*

The validation mechanism aims to detect if the object is correctly placed in the target position. This information is critical to enable a correct control of the transition between assembly stages. This process uses only depth information captured in real-time, because the projected light used in the second method would interfere with the color imagery. It uses a template matching approach to compare two images generated from 3D data taken by a RGB-D camera in the same perspective and thus validate if the piece is in the correct pose. 3D data is processed to extract only objects above the table and since the camera's view direction is perpendicular to the table, we obtained a representative 2D image from the top face of the pieces, which are used as templates afterwards.

Figure 4 presents the workflow of the validation process, which starts by capturing a point cloud using the camera (mounted in such a way as to detect the face of a table as presented in Figure 2) and send it to a computer for processing. Subsequently, a dedicated software running on the computer filters the objects outside a pre-established working area and segments the existing pieces on top of the table, to extract the associated clusters of points. Afterwards, the template of the current assembly stage is loaded and compared to the obtained cluster, checking if a match is found. Then, a decision has to be made (**decision 1**) verifying if the point cloud density is equal. In affirmative case, the process advances to **decision 2**. Otherwise, the process moves to **decision 3**. Later, in decision 2 the process verifies if the piece is in the correct pose. To perform this verification an image corresponding to the face of each piece facing the camera is compared, without texture, against the face of the pieces on the table. If they overlap above a threshold that was empirically determined (88%), the algorithm advances to **decision 4**. Otherwise, the process moves to **decision 3**. When **decision 3** occurs, the process verifies if it is the last cluster on the table. If so, the process moves to capture a new point cloud. Else, the process moves to match the cluster with the template. Finally, in **decision 4** the process verifies if the assembly is complete and either informs the user through the AR-based user interface or presents the next instruction and repeats the validation process for a new piece.

At the end of the overall validation process, the prototype provides AR-based visualization of the time taken to accomplish each piece alignment with the corresponding piece.

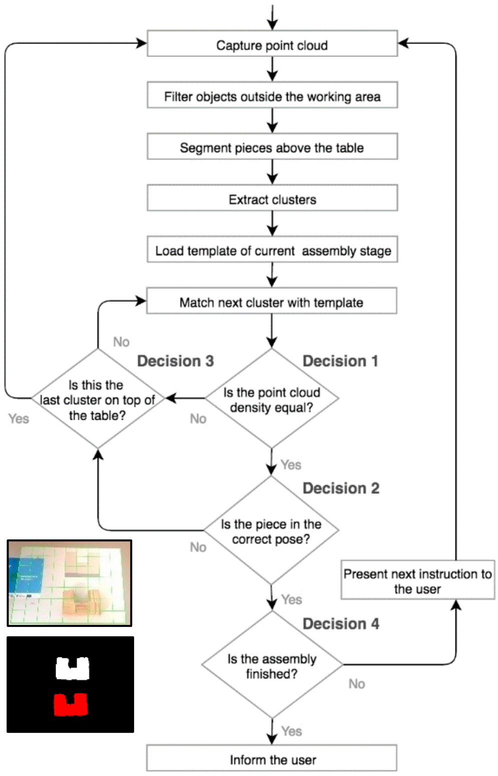

Figure 4 – Workflow of the validation process (used in both methods) through object recognition. The two images regarding the 'No' step – Decision 2 (verification of piece pose) show the real world in the top image, and the corresponding template matching procedure in the bottom image. This last image shows the desired and current pose of the piece in white and red respectively.

IV. The User Study

This study aimed to compare the usability and acceptance of the two AR-based assembly guidance methods. After an exploratory user test three experimental conditions were considered since many users complained about holding the mobile devices and thus a mobile-tripod configuration was also considered.

A. Experimental Design

A within-group experimental design was used. The null hypothesis (H0) considered was that the three experimental conditions are equally usable to mount a pre-defined pentomino puzzle.

The independent variable was the information display method provided to the users, with three levels corresponding to the experimental conditions: **1- Mobile AR using a handle (C1)**: The user performed the assembly using visual instructions presented on the screen of a mobile device, while using a handle (Figure 5 - Left). **2- Mobile AR using a tripod (C2)**: The user performed the assembly using visual instructions, also presented on the screen of a mobile device, but this time the mobile device was on a tripod in order to free both hands (Figure 5 - Middle). In both conditions, the instructions are overlaid on a live stream captured by the mobile device camera. **3- Spatial AR using a Kinect and a Projector (C3)**: The user performed the assembly using visual instructions presented by the projector (Figure 5- Right).

Performance measures including task performance, perceived mental workload, and participants' opinion were the dependent variables. The order in which the conditions were used, as well as participants' demographic data and previous experience with AR and assembly were registered as secondary variables. To minimize learning effects during the experiment, the participants were split into three groups and each group performed the three conditions in different orders.

Figure 5 – Study conditions: Summary of the data for all conditions. C1 – Mobile AR using a handle (Left); C2 – Mobile AR using a support (Middle); C3 – Spatial AR (Right).

B. Task

The participants were required to complete an assembly process of a Pentamino puzzle according to a set of instructions presented using the three experimental conditions. Figure 6 shows all existing pieces and the given name of each one, which are the result of 5 squares concatenated by the edges. For this study, we selected only 5 pieces, representing a generic assembly process (applicable to general assembly tasks rather than assembly tasks in specific domains) with a known solution (Figure 6), aiming to minimize bias towards a population with expertise in a certain domain possibly related to an assembly process.

The assembly process consisted of 5 procedural instructions that were 3 dimensional in nature. During each step, participants were required to place the piece in a pre-defined position (in a specific position and orientation), after selecting a piece of a specific size and color from a set of randomly distributed pieces on the workbench. In all conditions, virtual objects were aligned with the intended position of the desired piece on top of a virtual grid (Figure 7). For each experimental condition participants had to complete a puzzle slightly different whereas having similar complexity.

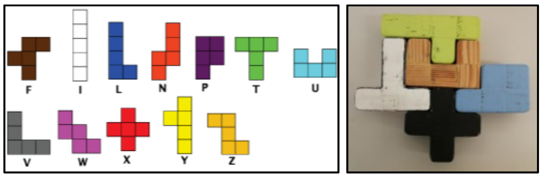

Figure 6 – Pentomino pieces identified by letters (Left). Pentomino Puzzle – final assembly (Right).

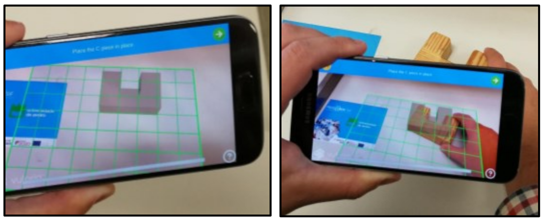

Figure 7 – Virtual grid highlighting the intended position and orientation of a 'U' shaped Pentomino piece.

C. Measurements

The data collection was conducted under the guidelines of the Declaration of Helsinki. Three types of measurements were taken: task performance, perceived mental workload and participants' opinion. Task performance is defined based on the time of completion (logged by the device, measured in seconds) and number of errors. Time of completion is the period needed to complete all procedures with each condition. Number of errors is the measurement of the number of inaccurate or incorrect actions performed by the participants during each task, where an error can be defined as a piece with the wrong color is inserted (E1), a piece inserted at the wrong location (E2) or a piece inserted with the wrong orientation (E3).

Mental workload (6 categories: mental demand, physical demand, temporal demand, effort, performance, frustration level) perceived by the participants during the experiment was measured using the NASA Task Load Index (NASA TLX) [12], based on their experience regarding the tasks performed using a 100-point scale. A mean weighted workload score is calculated by adding up the rating multiplied by an appropriate weighting for each category.

Participants' opinion was obtained through a post-experience questionnaire (taking into account the works by

[3][13][14][15]), including: demographic information (age, gender, previous experience with Pentomino, VR, AR, and AR in assembly tasks), the NASA TLX rating, and questions concerning the three conditions in order to assess the performance and ease of use of each condition, as well as which was preferred (Table 1).

Table 1 - Post-task Questionnaire for evaluating the performance and ease of use of the three conditions.

#	Question	Response Type
Q1	Was the amount of information displayed appropriate?	Likert Scale (1:Very Low; 7:Very High)
Q2	Were the assembly instructions difficult to understand?	Likert Scale (1:Very Low; 7:Very High)
Q3	The need to change attention between the mobile device and the pieces cause loss of focus on the objective?	Yes/No
Q4	Order the methods according to the preference.	Open Answer
Q5	What were the main difficulties?	Open Answer
Q6	After the experiment, I had symptoms of:	Multiple Choice
Q7	Add any suggestions you may consider relevant.	Open Answer
Q8	Add additional comments you may consider relevant.	Open Answer

D. Procedure

All participants used the 3 experimental conditions, but the order was varied among participants to avoid bias due to learning effects. At the beginning of the experiment, participants were instructed about the experimental setup, the tasks and gave their informed consent. Then they were asked to consider two levels of priorities: perform the task as accurately and as fast as possible. Afterwards, participants completed the assembly task and were observed by an experimenter who assisted them if they asked for help and used a standard form to make annotations (e.g. the number and type of errors, etc.). Immediately after completing the task using the three conditions, participants answered the post-task questionnaire.

E. Participants

Fifteen participants (5 female) aged from 23 to 43 years old, performed the assembly task and completed the post-experience questionnaire afterwards. Participants had various professions (e.g. Master and PhD students, Researchers, Teachers and a Geological Engineer). 9 participants had never assembled a pentomino puzzle, 2 had never used AR, and 12 had never used AR for assembly guidance purposes before.

V. Results and Discussion

This section presents and discusses the main results obtained from performance measures and opinion. Table 2 presents a summary of the participants' data for all conditions.

A. Completion time in each condition

Figure 8 shows the boxplot of the time required to complete the pentomino assembly for each experimental condition. Participants were faster when they used the third condition (Spatial AR – C3) (average=36.62s).

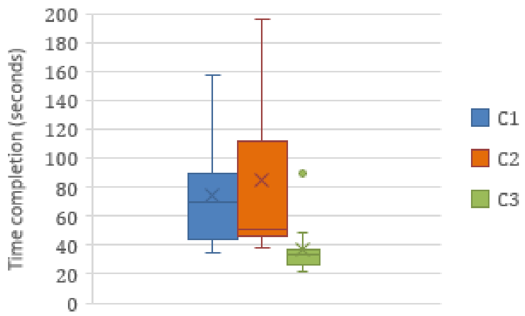

Figure 8 –Time of completion in each condition. C1 – Mobile AR using a handle; C2 – Mobile AR using a tripod; C3 – Spatial AR.

B. Total number of errors in each condition

The total number of participants' errors for each experimental condition is presented in table 2. Participants made less errors when they used the third condition (Spatial AR – C3) (average: E1=1, E2=1, E3=0).

C. Average score of NASA TLX rating

Figure 9 shows the boxplot of the NASA TLX rating for each experimental condition. Results show that participants had a slightly lower mental workload in the third condition (Spatial AR – C3) (20 out of 100), whereas subjects in the second condition (Mobile AR with tripod – C3) (24 out of 100) had a slightly higher mental workload.

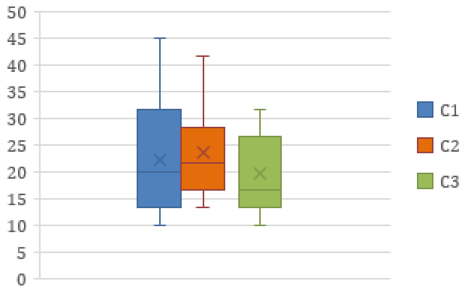

Figure 9 – NASA TLX scores (Task Load Index) in each condition. C1 – Mobile AR using a handle; C2 – Mobile AR using a tripod; C3 – Spatial AR

Table 2 – Summary of the data for all conditions. C1 – Mobile AR using a handle; C2 – Mobile AR using a tripod; C3 – Spatial AR. Error types: E1 – a piece with the wrong color is inserted, E2 – a piece is inserted at the wrong location or E3 – a piece is inserted with the wrong orientation. Each type refers to the total number of errors made in each condition.

Condition	Average Time (seconds)	Errors			NASA TLX rating
		E1	E2	E3	
C1	74.11s	1	2	2	22/100
C2	85.32s	0	4	3	24/100
C3	36.62s	1	1	0	20/100

D. Post-Experience Questionnaire

Answers to the post-task questionnaire show that the majority of the participants thought that the amount of information displayed was appropriate and the assembly instructions were easy to understand. Besides, none of the participants reported any symptoms of nausea, sickness and/or disorientation. In addition, the majority of the participants (13 out of 15) preferred the Spatial AR, followed by the Mobile AR with a tripod and last the mobile AR with a handle.

In the second condition, the initial goal of using a tripod to free participants (both) hands for the assembly process proved to be misleading and confusing. Participants felt this condition was mentally and physically more demanding, since the field of view provided by the device was static and thus limited the user movements when placing a piece. Additionally, some participants (4 out of 15) felt the need to pick-up the tripod, whenever localization and/or orientation errors occurred, with the goal of forcing a new (more convenient) field of view to accomplish the assembly process. Regarding the use of mobile AR, 8 out of 15 participants do not believe the need to change attention between the mobile device and the pieces caused loss of focus on the task objective.

Concerning future improvements, participants identified the following AR enhancements: display of lines, circular arrows and/or other hints, to guide participants to manipulate the pieces into the correct poses, thus recovering from localization and/or orientation errors. Likewise, participants also recognized that the AR methods could be more beneficial when applied to more complex tasks. Thus, in a future study it is important to address complexity of assembling tasks [16]. Similarly, a new study including more complex tasks and a larger group of participants may obtain significant differences, providing more insight concerning the proposed methods.

Some users (2 out of 15) complained about the accuracy required to advance to the next step and the lack of feedback from the application. These complaints happened when the pieces were misaligned, and the users did not know what they should do. In the second situation three users reported that the perspective effect associated with the transparency of the pieces led them to think that one of the pieces was above another and not on the table top.

VI. Conclusions and Future Work

Augmented Reality (AR) has been considered as having great potential for assisting in assembly tasks. This paper presented a prototype using two different AR-based methods (two based on mobile AR and another using Spatial AR) with validation using object recognition, to provide assembly guidance. A controlled experiment was performed to compare the performance, ease of use and acceptance of the methods. Participants were faster and made slightly fewer errors using the Spatial AR condition. Besides, participants preferred this condition. In addition, Nasa TLX rating showed that this condition has a slightly lower cognitive load on the participants.

Work will continue through the definition of more complex tasks, improvement of the object recognition algorithm to enable tracking of 3D pieces (possibly using machine learning mechanisms), and presentation of situated visualizations (visualization in context) to better guide assembly tasks. With this, we aim at making the prototype more suitable for large scale user-testing, and thus obtaining more insight concerning the proposed methods.

References

[1] L. C. Wu, I. C. Lin, and M. H. Tsai, “Augmented reality instruction for object assembly based on markerless tracking", ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, I3D ’16, pp. 95–102, 2016.

[2] R. T. Azuma, “A Survey of Augmented Reality”, Presence: Teleoperators and Virtual Environments, vol. 6, no. 4, pp. 355–385, 1997.

[3] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. La Viola, T. Mashita, and H. Takemura, “The effectiveness of an AR-based context-aware assembly support system in object assembly”, IEEE Virtual Reality, pp. 57–62, 2014.

[4] S. J. Henderson and S. K. Feiner, “Augmented reality in the psychomotor phase of a procedural task”, in IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, pp. 191–200, 2011.

[5] A. Tang, C. Owen, F. Biocca, and W. Mou, “Comparative Effectiveness of Augmented Reality in Object Assembly”, Conference on Human factors in computing systems, CHI '03, pp. 73–80, 2003.

[6] A. Nishihara and J. Okamoto, “Object recognition in assembly assisted by augmented reality system”, in SAI Intelligent Systems Conference, IntelliSys 2015, pp. 400–407, 2015.

[7] F. Loch, F. Quint, and I. Brishtel, “Comparing video and augmented reality assistance in manual assembly”, Conference on Intelligent Environments, IE 2016, pp. 147–150, 2016.

[8] Di. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui, “Mobile Augmented Reality Survey: From Where We Are to Where We Go”, IEEE Access, vol. 5, pp. 6917–6950, 2017.

[9] F. Zhou, H. Duh, and M. Billinghurst, “Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR”, IEEE International Symposium on Symposium on Mixed and Augmented Reality, ISMAR 2008, pp. 193–202, 2008.

[10] K. Kim et al., “Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008 – 2017)”, IEEE Transactions on Visualization and Computer Graphics, 2018.

[11] R. T. Azuma, “The Most Important Challenge Facing Augmented Reality”, Presence, vol. 25, no. 3, pp. 234–238, 2016.

[12] S. G. Hart, “Background Description and Application of the NASA Task Load Index (TLX)”, Department of Defense Human Engineering Technical Advisory Group Workshop on Workload, 1986.

[13] Duenser, A., Dünser, A., & Billinghurst, M. (2011). Evaluating Augmented Reality Systems. In Handbook of Augmented Reality (pp. 289–307).

[14] M. De Paiva Guimarães and V. F. Martins, “A checklist to evaluate augmented reality applications”, Symposium on Virtual and Augmented Reality, SVR 2014, pp. 45–52, 2014.

[15] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. S. Ii, “A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014”, Frontiers in Robotics and AI, vol. 5, no. April, 2018.

[16] A. Syberfeldt, O. Danielsson, M. Holm, and L. Wang, “Visual Assembling Guidance Using Augmented Reality”, Procedia Manufacturing, vol. 1, pp. 98–109, 2015.