

Article

# User Interactions for Augmented Reality Smart Glasses: A Comparative Evaluation of Visual Contexts and Interaction Gestures

Minseok Kim, Sung Ho Choi, Kyeong-Beom Park and Jae Yeol Lee \*

Department of Industrial Engineering, Chonnam National University, Gwangju 61186, Korea

\* Correspondence: jaeyeol@chonnam.ac.kr; Tel.: +82-62-530-1782

Received: 25 June 2019; Accepted: 1 August 2019; Published: 4 August 2019



**Abstract:** Smart glasses for wearable augmented reality (AR) are widely used in various applications, such as training and task assistance. However, as the field of view (FOV) in the current AR smart glasses is narrow, it is difficult to visualize all the information on the AR display. Besides, only simple interactions are supported. This paper presents a comparative and substantial evaluation of user interactions for wearable AR concerning visual contexts and gesture interactions using AR smart glasses. Based on the evaluation, it suggests new guidelines for visual augmentation focused on task assistance. Three different types of visual contexts for wearable AR were implemented and evaluated: stereo rendering and direct augmentation, and non-stereo rendering and indirect augmentation with/without video background. Also, gesture interactions, such as multi-touch interaction and hand gesture-based interaction, were implemented and evaluated. We performed quantitative and qualitative analyses, including performance measurement and questionnaire evaluation. The experimental assessment proves that both FOV and visual registration between virtual and physical artifacts are important, and they can complement each other. Hand gesture-based interaction can be more intuitive and useful. Therefore, by analyzing the advantages and disadvantages of the visual context and gesture interaction in wearable AR, this study suggests more effective and user-centric guidance for task assistance.

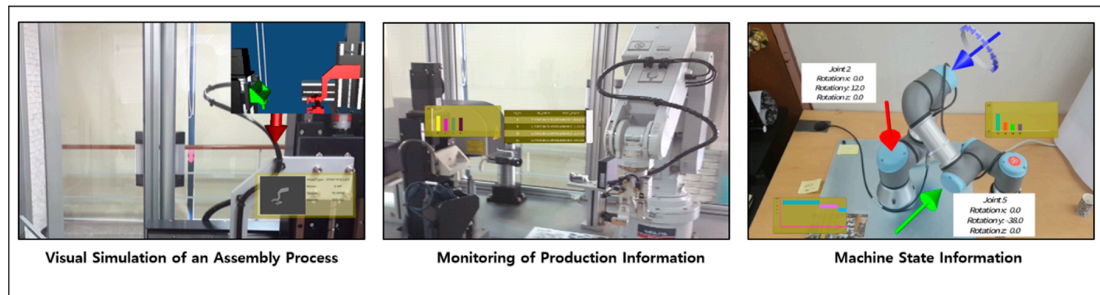
**Keywords:** interaction gesture; smart glasses; task assistance; user interaction; visual context; wearable AR

## 1. Introduction

Augmented reality (AR) has been widely utilized in various fields because it can visually superimpose different information, such as text, images, videos, and 3D virtual objects, onto corresponding physical artifacts. In the early stage of AR, it was mainly applied to entertainment, such as games, but recently it has been expanded and applied in various other fields, such as assembly, maintenance, training, remote collaboration, and smart manufacturing [1–5].

In different manufacturing fields, an industrial worker can perform a variety of tasks, such as monitoring, maintenance, assembly, and inspection of equipment and products [6]. In performing these tasks, the worker must review the dynamically changing status of equipment and products in real-time and take appropriate actions [7]. AR can support task assistance by overlaying machine status information onto a related physical artifact and by providing step-by-step instructions. Thus, the worker can perform their tasks more effectively. However, the worker must use a separate handheld device, such as a smartphone or tablet, to acquire task assistance. In particular, novice workers are required to have other training and instruction manuals when performing such tasks in an actual manufacturing environment. Thus, the worker cannot observe the equipment, products, and handheld device at the same time while performing tasks because their attention would be distracted.

For this reason, wearable AR with smart glasses is considered to help industrial workers to more effectively perform their tasks [7–9]. Different types of AR smart glasses have been developed and applied to various industrial fields. Figure 1 shows wearable AR-based task assistance to workers wearing AR smart glasses, such as visualizing the simulation of assembly, monitoring production information, and inspecting the status information of a robot.



**Figure 1.** Visual and interactive assistance through wearable augmented reality (AR) applied to manufacturing services, such as visualization of the status of machines and robot.

AR smart glasses, such as Epson Moverio [10] and Microsoft HoloLens [11], have been developed and commercialized to provide users with wearable AR applications in mobile environments. In the wearable AR, it is essential to help the worker to understand and interact with the AR-based visual instructions more intuitively with a less cognitive burden. For example, when the worker must follow different types of annotations augmented on the display of the smart glasses (right of Figure 1), it is necessary to provide user-friendly visual augmentation forms on the optical see-through display of the AR smart glasses.

However, AR smart glasses with a see-through display have a limited display size and viewing angle [6]. For example, Epson Moverio BT-200 has a 0.42 inch wide LCD panel, a dual display with  $960 \times 540$  resolution, and a field of view (FOV) of approximately 23 degrees [10]. The FOV in current smart glasses is not large enough to display all the necessary information, which limits user-centric task assistance. Furthermore, the method for user interactions in the wearable AR environment is not natural and limited not to be applied to manufacturing services since it cannot support two-handed operations, and interaction is only supported by a multi-touchpad (e.g., Epson Moverio). Recently, two-handed operations have been supported in other smart glasses, such as MS HoloLens, but it can only support simple interactions, such as clicking and selection.

It is crucial to analyze user interactions concerning not only visual contexts but also interaction metaphors to help the worker to effectively utilize AR smart glasses with limited capabilities for task assistance [12]. Note that usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. On the other hand, user experience is a person's perceptions and responses that result from the use or anticipated use of a product, system or service, and thus it includes all the users' emotions, beliefs, preferences, perceptions, etc. [12].

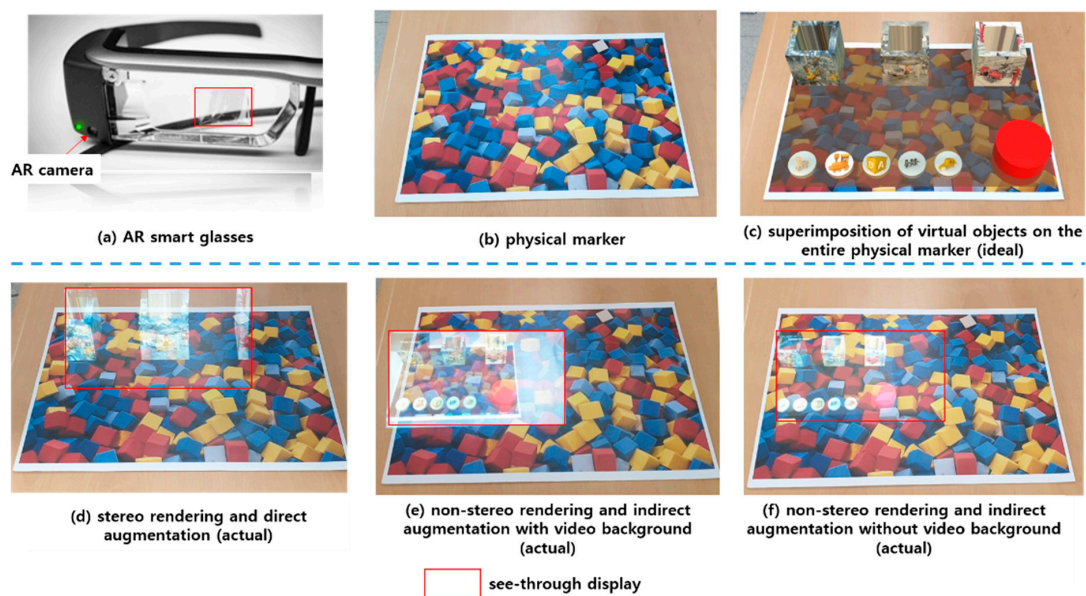
Therefore, this paper proposes a comparative and substantial evaluation of user interactions for wearable AR concerning the visual context and interaction gesture using AR smart glasses. It also suggests new guidelines for visual augmentation focused on task assistance. We performed quantitative and qualitative analyses, including performance measurement and questionnaire evaluation. By analyzing the advantages and disadvantages of wearable AR for visualization and interaction, we could provide more effective guidance for user-centric task assistance.

In this research, the visual augmentation, according to the visual context in the see-through display of AR smart glasses, is classified as follows.

## 1. Stereo rendering and direct augmentation

2. Non-stereo rendering and indirect augmentation with video background
3. Non-stereo rendering and indirect augmentation without video background

The stereo rendering and direct augmentation method superimposes virtual information directly onto its physical artifact through the see-through display of smart glasses (Figures 2d and 3a). The stereo rendering can provide stereo rendering-based immersive visualization, and the overlaid virtual information is almost matched or registered with its corresponding physical artifact. For this reason, a registration process is required to match the virtual information with the physical artifact. The non-stereo rendering and indirect augmentation with video background overlays virtual information indirectly onto the physical artifact in the image taken by the smart glasses' AR camera and then provides the augmented image to the user through the see-through display of AR smart glasses (Figures 2e and 3b). In this case, the virtual information does not match with the physical artifact concerning the user's viewpoint. The non-stereo rendering and indirect augmentation without video background method removes the camera image from the non-stereo rendering and indirect augmentation with video background and presents only the virtual information on the display (Figures 2f and 3c). This method also does not support matching virtual information with the physical artifact.



**Figure 2.** Visual augmentation concerning visual contexts in the real AR smart glasses.

For most AR smart glasses, such as Epson Moverio (Figure 2a), the stereo rendering and direct augmentation approach has narrower FOV than the non-stereo rendering and indirect augmentation through the see-through display. Instead, it can provide consistent visual augmentation with physical objects. Therefore, different visual contexts have their advantage and disadvantage concerning FOV and visual registration, which can influence task performance and usability. However, there is little research to evaluate this issue substantially.

Besides, since most AR smart glasses can only support simple interactions using the touch gesture and hand gesture, it is challenging to provide more user-centric interactions using full hand gestures. However, one drawback is to attach another sensor, such as a depth sensor to the smart AR glasses, which may prevent the worker from conducting tasks and cause another cognitive load. Besides, it is essential to calibrate the coordinate of wearable AR and that of the hand tracking sensor. Furthermore, it is still necessary to evaluate the usability of interactions concerning task performance and the cognitive load of the industrial worker in an actual manufacturing environment.

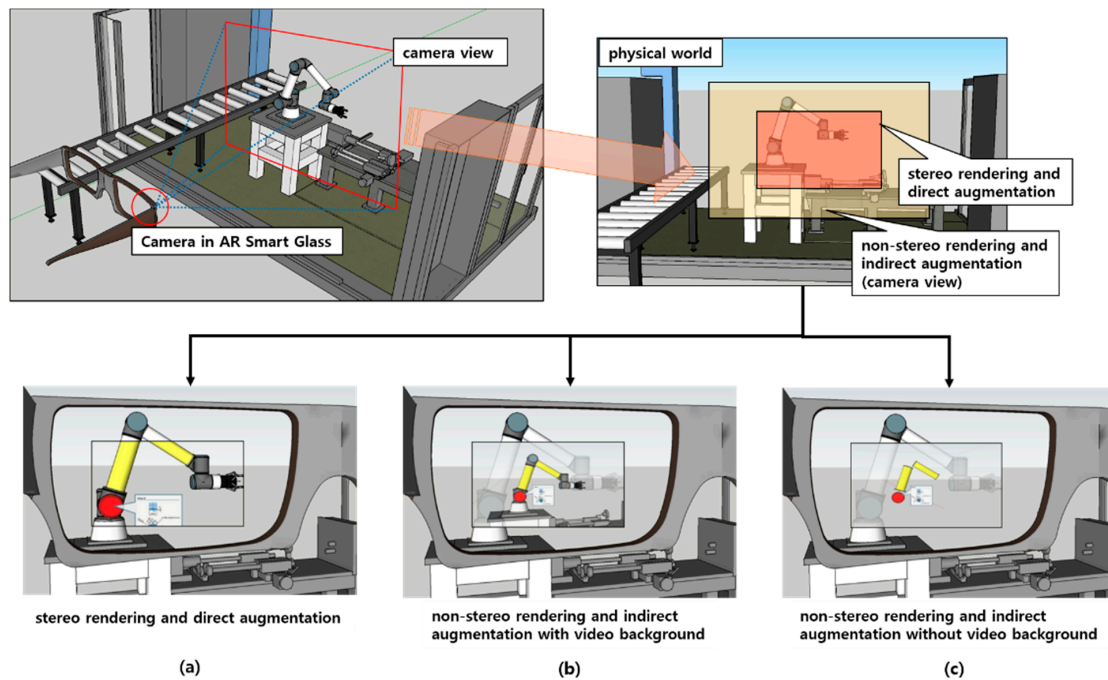


Figure 3. Three different types of visual contexts for AR smart glasses.

As the most popular gesture interactions in wearable AR are touch-based and gesture-based [6], we have classified them as follows.

- Hand gesture-based interaction
- Multi-touch interaction

Through the hand gesture-based interaction, the user wearing AR smart glasses can manipulate virtual objects through hand gestures directly (Figure 4a). Besides, it supports bi-manual operations as the worker can use both hands. However, to support hand gesture-based interaction, an additional depth sensor attached to the front of smart glasses is required, and coordinate systems between the smart glasses and the depth sensor must be calibrated. On the other hand, the multi-touch interaction uses a touchpad connected to the AR smart glasses (Figure 4b). The usage is similar to the touch interface of a smart device, such as the smartphone or tablet. Thus, its interaction is familiar to the worker. However, it cannot support bi-manual operations as the worker must touch the pad with one hand.

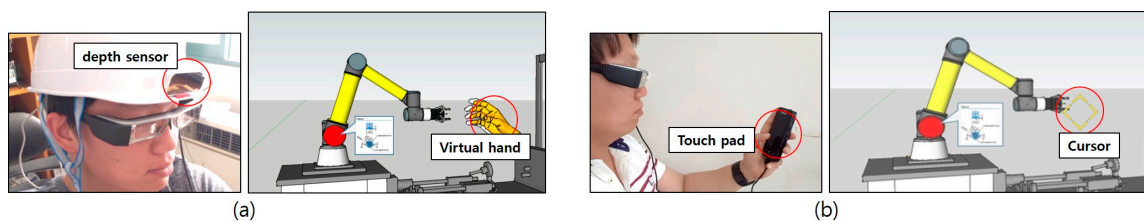


Figure 4. Two different types of AR interactions: (a) hand gesture; (b) touch gesture.

This study has the following contributions.

- We have conducted a comparative and substantial evaluation of user interactions for AR smart glasses concerning visual contexts and gesture interactions, which suggests new directions for task assistance in wearable AR.
- We have performed quantitative and qualitative experiments to analyze task performance and usability to provide user-centric visualization and interaction.

- Through the experiment, we have found that FOV and visual registration complement each other so that it is necessary to increase the visual focus of the information by calculating degree-of-interest.
- The analysis result was tested by a real use case of industrial workers concerning cognitive load by performing tasks using wearable AR.

This paper is organized as follows. Section 2 presents related work. Section 3 defines the visualization and interaction methods in wearable AR. Section 4 describes the first user study to compare the methods for visual contexts and gesture interactions through virtual object manipulations. Experimental results were evaluated through quantitative and qualitative analyses. Section 5 presents the second user study to assess the methods by comparing different FOVs and evaluating a 3D matching operation between physical and virtual objects. Section 6 describes a real use case of industrial workers who have experienced the proposed approach. Section 7 discusses user studies for user-centric task assistance with AR smart glasses and suggests new challenges. Section 8 concludes the paper and suggests future works.

## 2. Related Work

In recent years, wearable AR using smart glasses has been applied to various applications and evaluated for user-centric task assistance [1,8,13–15]. Previous research works in wearable AR can be classified into two areas: (1) visual context-based visualization and (2) gesture interactions in the physical environment.

### 2.1. Visual Context-Based Visualization

Several research works have been conducted to provide visualization methods concerning visual contexts for practical task assistance in different environments [1,8,9,13]. Most smart AR glasses support see-through augmentation. However, regarding visual contexts, they can provide either direct or indirect augmentation on physical artifacts. Direct augmentation requires visual registration between physical and virtual objects. Although it can provide a more immersive feeling, its FOV is narrower than the indirect augmentation. Therefore, there is a trade-off between these augmentations concerning visual contexts. However, there is little research to analyze visual contexts in wearable AR for task assistance.

Previous studies proposed methods for task assistance in wearable AR. Zhu et al. [9] proposed AR-Mentor, which is an AR system for assisting in maintenance and repair tasks. Their method combined AR smart glasses with six degrees of freedom (DOF) pose tracking and virtual personal assistance to instruct a novice in performing maintenance tasks. Funk et al. [16] proposed HoloCollab, which is a system using AR smart glasses to create an augmented training experience for both trainer and trainee. Some other works have attempted to combine an optical see-through display with projector-based spatial augmented reality [17], and have discussed how a working system should be developed by integrating smart glasses with a question-answering module [18].

The visualization methods were also compared and evaluated in various devices [19,20], including tablets, head mounted displays (HMDs), and projectors. Zeng et al. [1] investigated what characteristics of eye-wearable technology impact user performance in machine maintenance by comparing eyewear-central and eyewear-peripheral visualizations depending on the position of the display. They found that the eyewear-central display resulted in faster task completion time than the eyewear-peripheral display. Besides, the influence of the graphical context on task performance was also evaluated [8,13,14]. Robertson et al. [13,14] assessed the effects of registration error in a Lego block placement task and the effectiveness of graphical context for task assistance. Their work showed that adding contexts to a scene due to registration error could help a user perform given tasks more effectively by reducing the number of mistakes and task completion time. Khuong et al. [8] evaluated the effectiveness of an AR-based context-aware assembly support system using two different visual

overlays. Their experimental results showed that the visual overlay that presented target status next to real objects outperformed the direct overlaid visualization on the actual objects.

## 2.2. Gesture Interactions

The user can perform either a 2D-based task, such as selecting buttons and typing texts, or a 3D-based task, such as 3D orientation and translation [21]. Therefore, offering AR smart glasses with better input metaphors makes the interaction more intuitive and efficient, which enables the user to handle more complicated and visually demanding tasks [15]. In particular, this paper describes two widely used approaches: gesture-based interaction and touch-based interaction.

Gesture-based interaction tracks the position and orientation of the user's fingers or hands using the camera or inertial measurement unit (IMU) of the AR smart glasses and supports interactions with virtual objects in the AR environments [21–30]. Ha et al. [21] proposed WeARHand, which allows the user to manipulate virtual 3D objects with a bare hand in a wearable AR environment. Dudley et al. [29] proposed a system that enabled users to type on a virtual keyboard using a head-mounted AR device and hand localization derived from body-fixed sensors. Meanwhile, Hsieh et al. [28] distilled a set of design principles for mid-air gestural interaction for smart glasses, one of which showed the potential of using an independent hand tracking device, such as a haptic glove, to form an integrated wearable system with smart glasses. Hand gesture-based interaction can support natural gestures, such as pointing and grabbing, to interact with virtual objects [15].

In other research works, touch-based interaction has been widely used in various smart and mobile devices [31–36]. Ahn et al. [33] proposed using a smartwatch as an indirect input device to make an effective smart glasses interaction method for tasks, such as text entry. Whitmire et al. [34] proposed DigiTouch, which is a touch-sensitive glove that enables thumb-to-finger interaction for eyes-free input on wearable systems. DigiTouch uses thin, partially conductive fabric strips along with the fingers and a conductive patch on the thumb pad. Wang et al. [35] presented PalmType, which uses palms as interactive surfaces to enable an intuitive and efficient keyboard interface for smart glasses. Lissermann et al. [36] proposed the augmentation of accessories worn behind the ear with a device that unobtrusively instrumented the ear as an interactive surface.

Some researchers have also evaluated interaction methods using AR smart glasses [37,38]. Terrenghi et al. [38] found that the modes of manipulation of physical versus digital media were fundamentally different. However, previous methods have not been fully evaluated by regarding both the visual context and the interaction gesture, although both aspects may affect each other.

Although previous research works presented various approaches to the visualization and interaction in wearable AR environments, most of them have focused on developing domain-specific applications rather than conducting comparative evaluations for user-centric task assistance concerning different types of visual augmentation. Regarding the visual context, there has been little research on evaluating task performance through the comparison between FOV and visual registration in wearable AR. In terms of the gesture interaction, further study is still needed for task assistance with combined gesture interactions in manufacturing operations. Note that some operations were effective in 2D-based tasks, such as selecting buttons, but others were useful for 3D-based tasks, such as 3D rotation and translation of virtual objects [15].

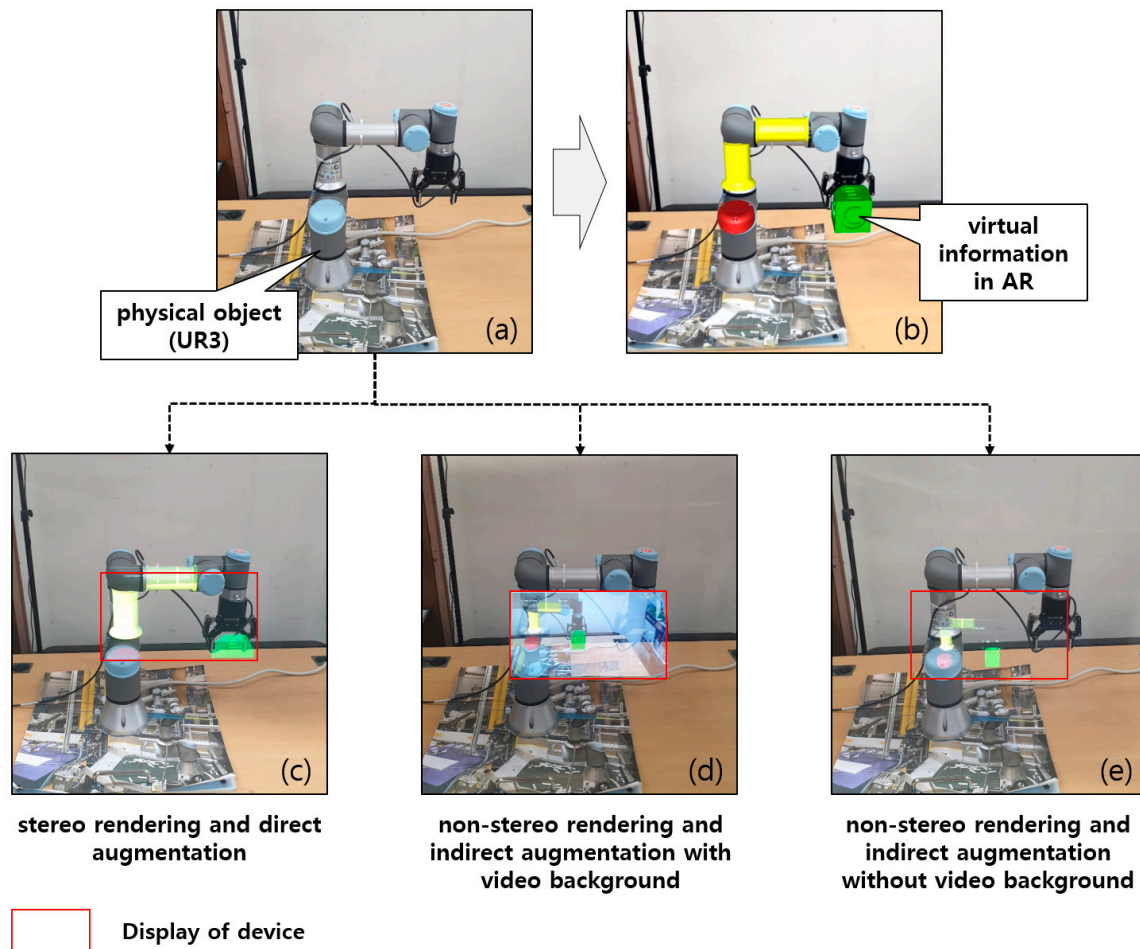
Furthermore, although several research works have suggested new interactions, there is further room for improvement in wearable AR. Therefore, in this study, we analyzed not only the usability of the visual context but also interaction methods using AR smart glasses. Quantitative and qualitative analyses were carried out by comparing and evaluating task performance. This paper also presents a new challenge by analyzing experimental results.

## 3. Visual Context and Gesture Interaction

This section defines visual contexts and gesture interactions for comparative evaluation in wearable AR.

### 3.1. Visual Augmentation Depending on Visual Contexts

We classified three different types of visual augmentation depending on how virtual objects are superimposed on the optical see-through display of AR smart glasses. As shown in Figure 5, there are two differences in the visual augmentation, depending on the visual context. One difference is the amount of information visualized in the display of smart glasses. The other is how exactly the physical artifact and its corresponding virtual object are matched through the display or the camera.



**Figure 5.** Visual augmentation depending on visual context in a real environment. UR3: Universal Robot-3.

The stereo rendering and direct augmentation is a visualization method that supports almost seamless matching between physical and virtual objects through the see-through display. Thus, a high degree of immersion can be provided to the user, and both the physical object and the virtual object can be viewed without mismatch. However, as the FOV of the see-through display for human eyes is narrower, as shown in Figure 5, even if the virtual object can be matched with the physical object, the amount of information that can be visualized at once is smaller than that of the non-stereo rendering and indirect augmentation method. On the other hand, the non-stereo rendering and indirect augmentation with the video background is a method of augmenting a virtual object on a captured camera image that is rendered on the see-through display. Usually, the camera FOV is larger than that of the display of the AR smart glasses. However, the visual AR information does not match with a physical object. The matching is accomplished from the camera view, not from the user’s perspective. This visual augmentation provides more information to the user simultaneously even if accurate matching between the physical and the virtual objects is not achieved through the display of the smart glasses. However, inconsistency between the physical and virtual objects can cause a cognitive burden

on the user in the process of performing tasks. The non-stereo rendering and indirect augmentation without video background deletes camera background image in the second method and renders only virtual information. Thus, unlike the second method, the background image of the camera cannot be seen.

### 3.2. Gesture Interaction

Considering currently available AR smart glasses, interactions with virtual objects in the wearable AR environment can be classified into hand gesture-based interaction and multi-touch interaction, using a touchpad (Figure 4). In particular, as shown in Figure 6a, the hand gesture interaction is similar to how the user performs a task using a hand in a real environment. Through different hand gestures, as shown in Figure 6b, it is possible to achieve interaction, such as selecting (picking or grabbing), moving (dragging), zooming (pinching), or rotating a virtual object, that is visualized in wearable AR.



**Figure 6.** Hand gesture-based interaction (a) in wearable AR for monitoring a manufacturing service; (b) different types of hand gesture interactions.

To support hand gesture-based interactions with virtual objects, the position and posture of the user's hand must be consistently tracked and recognized. Furthermore, the traced real hand and the virtual hand in wearable AR must be synchronized on the display of AR smart glasses. For this purpose, we used a depth sensor, which tracks the real hand and acquires the positions and postures of the palm and finger joints in real-time. By matching the generated virtual hand with the actual hand, the user can directly and intuitively interact with the virtual object in the wearable AR space.

On the other hand, multi-touch interaction is easily performed on a touchpad connected to smart glasses, as shown in Figure 7. The touch interaction is very intuitive and natural. However, only one hand is available during the interaction. First, the user sees a cursor shown on the display of the smart glasses and moves the cursor on the touchpad to select the object. Then, they can interact with the virtual object through rotation and translation. However, as it is not possible to perform multi DOF interaction, such as simultaneous translation and rotation like in the hand gesture method, it is necessary to selectively and sequentially perform the interaction after switching from one mode to the other through a menu selection. However, during the communication in wearable AR, the touchpad must be held in one hand, which limits the use of both hands.





Figure 7. Multi-touch interaction using a touchpad in AR smart glasses.

#### 4. Comparative Evaluation of User Interaction in Wearable AR

##### 4.1. Overview of the Experiment

We conducted a comparative evaluation of user interactions concerning the visual context and gesture interaction. We assumed that FOV and visual registration would be influential factors for supporting user-centric task assistance, and the hand gesture would also provide more natural and intuitive interactions in wearable AR. The hand gesture is more natural in the physical environment, such as a manufacturing environment, and it can also support bi-manual operations. For evaluating gesture interactions, the visual context is based on the non-stereo rendering and indirect augmentation because most AR smart glasses can support non-stereo rendering and indirect augmentation rather than stereo rendering and direct augmentation [6,10,15].

##### 4.2. Experimental Setup

Epson's Moverio BT-200 was used as the AR smart glasses [6,10]. The transparent screen is located at the front of the user's view (eyewear-central), and a separate touchpad connected to the smart glasses is provided to enable multi-touch interaction. The wearable AR system was implemented using Unity3D [39]. We also used the Vuforia software development kit (SDK) [40] to track natural feature-based AR markers, which was used to overlay virtual objects onto the physical environment.

We used a Leap Motion sensor [41] as a depth sensor to track the user's actual hands to support the hand gesture interaction. As shown in Figure 8, the Leap Motion sensor was attached to the front of the safety helmet of the worker or the smart glasses, while the AR camera was embedded in the AR smart glasses. The absolute orientation algorithm [42,43] was used for the calibration of the two coordinate systems.

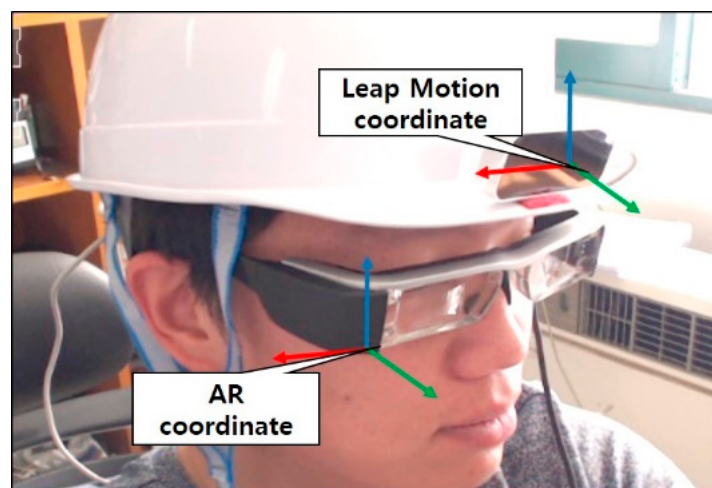


Figure 8. Difference between the wearable AR coordinate and the depth sensor coordinate.

In particular, as the origin, transformation, and scale are different between the coordinate of the camera of the smart glasses and that of the depth sensor, a calibration operation for matching the two coordinates must be performed, as shown in Figure 9.

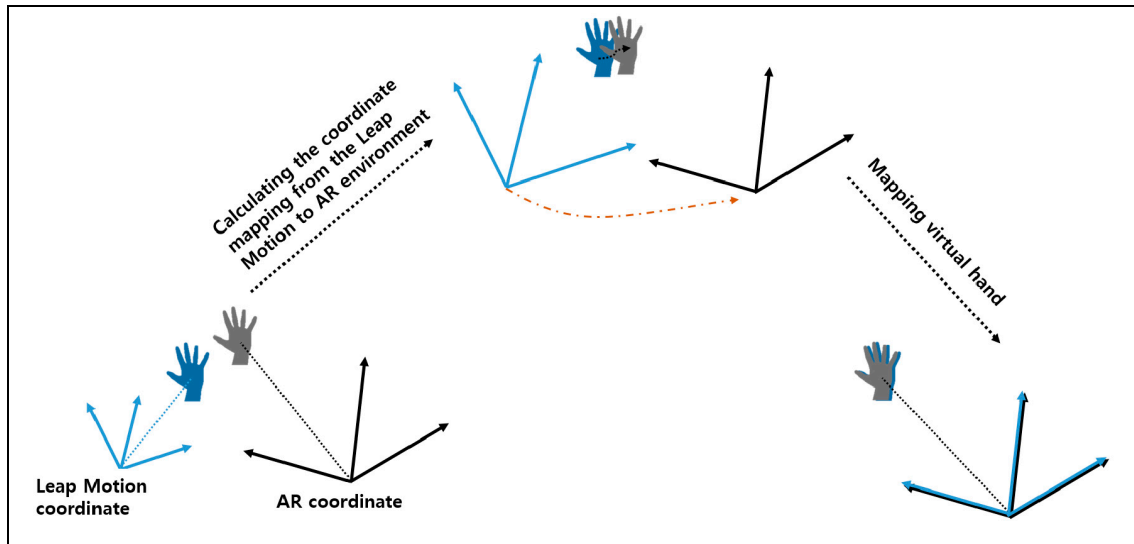


Figure 9. Calibration between the two coordinates.

Let  $X_{Leap} = \{x_1, x_2, \dots, x_n\}$ , a set of points in the Leap Motion coordinate, and  $Y_{AR} = \{y_1, y_2, \dots, y_n\}$ , a set of points in the wearable AR coordinate, be corresponding points in 3D space. The absolute orientation algorithm finds the similarity transformation parameters (rotation  $R$ , translation  $t$ , and scaling  $c$ ) that give the least mean squared error between these two point sets. The minimum value  $e^2$  of the mean squared error is defined as follows.

$$e^2(R, t, c) = \frac{1}{n} \sum_{i=1}^n \|y_i - (cRx_i + t)\|^2 \tag{1}$$

In summary, to minimize  $e^2$ , we have to find  $R, t, c$ , and such that

$$R = USV^T \tag{2}$$

$$c = \frac{1}{\sigma_{Leap}^2} trace(DS) \tag{3}$$

$$t = \mu_{AR} - cR\mu_{Leap} \tag{4}$$

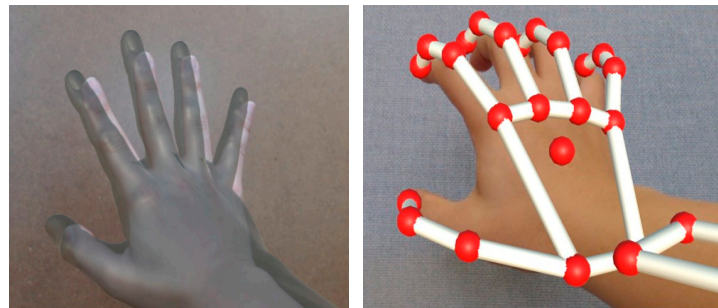
$$\sum xy = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_{AR})(x_i - \mu_{Leap})^T \tag{5}$$

where  $trace(DS)$  is the sum of the elements on the main diagonal of  $DS$ ,  $\sum xy$  is a covariance matrix of  $X_{Leap}$  and  $Y_{AR}$ ,  $\mu_{Leap}$  and  $\mu_{AR}$  are mean vectors of  $X_{Leap}$  and  $Y_{AR}$ ,  $\sigma_{Leap}^2$  and  $\sigma_{AR}^2$  are variances of  $X_{Leap}$  and  $Y_{AR}$ , respectively. Let a singular value decomposition of  $\sum xy$  be  $UDV^T$ ,  $D = \text{diag}(d_i)$ ,  $d_1 \geq d_2 \geq \dots \geq 0$ , and

$$S = \begin{cases} I & \text{if } \det(U)\det(V) = 1 \\ \text{diag}(1, 1, \dots, 1, -1) & \text{if } \det(U)\det(V) = -1 \end{cases} \tag{6}$$

The detailed explanation and the proof are described in Reference [43].

Through the calibration between the AR camera and the Leap Motion sensor, the real hand can be matched with a virtual hand in the wearable AR space, as shown in Figure 10.



**Figure 10.** Actual hand detection and calibration with the virtual hand in the wearable AR.

#### 4.3. Experimental Design

For the comparative evaluation of visual contexts and interaction gestures in the wearable AR environment, we designed two user studies. Three experimental tasks, such as (1) selection, (2) dragging and dropping, and (3) 3D matching, were designed to compare and evaluate the usability and task performance concerning the visual context and interaction gesture in wearable AR. Among the three tasks, two tasks, such as dragging and dropping and 3D matching, were conducted to evaluate the visual context. Two tasks, such as selecting and 3D matching, were conducted to assess the interaction gesture. When a subject looked at a specific area with a generic feature-based AR marker(s) through the camera of AR smart glasses, virtual objects were augmented on or above the AR marker, and then the participant was required to perform the given task.

Twenty university students (fourteen males and six females) were recruited with an average age of 25.9 (SD = 1.25) in a controlled laboratory environment [44]. Four of them had experience in AR applications, two had experience in using Leap Motion, and the others had no such experience. Before the experiment, subjects were instructed on how to perform the three tasks, along with a description of the visualization and interaction methods, and then practiced until they were almost familiar with each method. After practice, subjects performed full-scale tasks in a random order to avoid the learning effect. In other words, the order of conducting three tasks, and the order of user studies were randomly assigned to each subject.

Through the experiment, quantitative and qualitative assessments of visual contexts and interaction gestures were conducted. For the quantitative evaluation, the completion time of each task was measured. After the completion of all the tasks, surveys and interviews were also conducted for qualitative evaluation. The questionnaire consisted of questions that were evaluated based on a 7-point Likert scale (1: strongly disagree, 7: strongly agree). In particular, we used all the questions from the System Usability Scale (SUS) [45] without any modification. SUS is well-known as an objective questionnaire for usability evaluation. We have not listed them in this paper. Figure 11 shows a snapshot of the worker who performed a matching task with the Universal Robot-3 (UR3) robot. UR3 is a famous intelligent and collaborative robot widely used in manufacturing environments.



Figure 11. A worker was performing an experimental task.

### 4.3.1. Selection

Selection or picking is one of the fundamental tasks for the user interaction to perform a specific operation like pushing a button or an icon. Thus, we analyzed the task performance on the selection of virtual icons according to different visual contexts. Through this task, the subject can change the position and orientation of the robot directly. As shown in Figure 12, a virtual control panel for manipulating the UR3 robot is augmented on the see-through display of the AR smart glasses. Then, at the top of the virtual control panel, a required task is shown for the subject to select the icon either with a hand gesture or a touch gesture. The subject must understand the instruction and select the correct icon. For example, if ‘Force +’ or ‘Speed MIN’ are displayed, the subject has to select corresponding buttons on the see-through display. In the case of hand gesture-based interaction, the subject can choose the virtual icon directly with their finger. In the case of touch interaction, the user moves the cursor in the touchpad and selects the right button by touching the pad. This task continues until the subject selects all the required buttons to complete a required task.

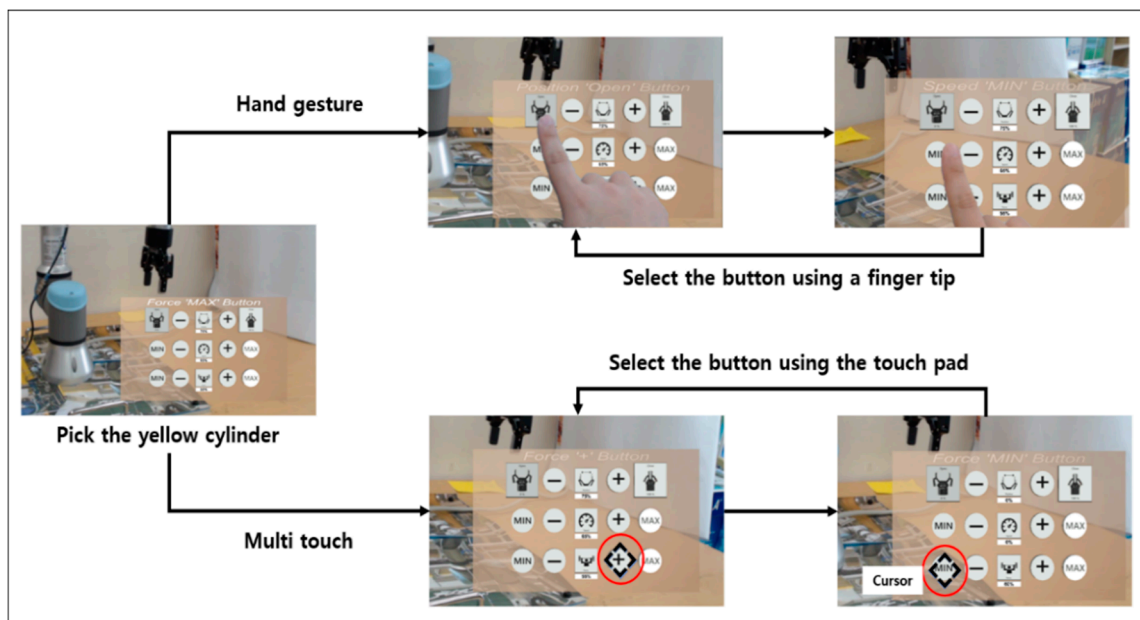
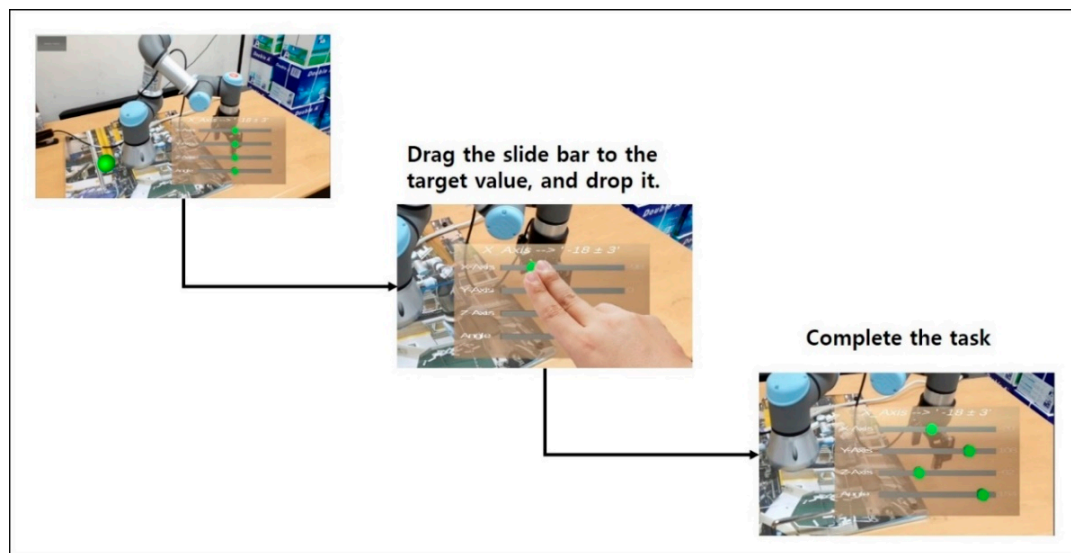


Figure 12. Selection task to operate the UR3 robot.

### 4.3.2. Dragging and Dropping

A detailed step for the dragging and dropping task is shown in Figure 13. The dragging and dropping task is a 2D operation on the 3D plane. Thus, this task evaluates the performance of 2D-based

manipulation in the 3D space. The mission of this task is to move a visual object to the desired position by dragging and dropping a specified virtual icon or slide bar in the virtual control panel. As shown in Figure 13, the virtual control panel displays the type of a slide bar and its attribute value to be adjusted. Then, the subject has to understand the required task and to slide the bar to the desired value using a hand gesture. For example, for the UR3 robot, the worker can pull a slide bar to change the joint angle of the robot or to change the speed of the robot movement. Finally, this task is completed when the subject controls the slide bar to the desired location.

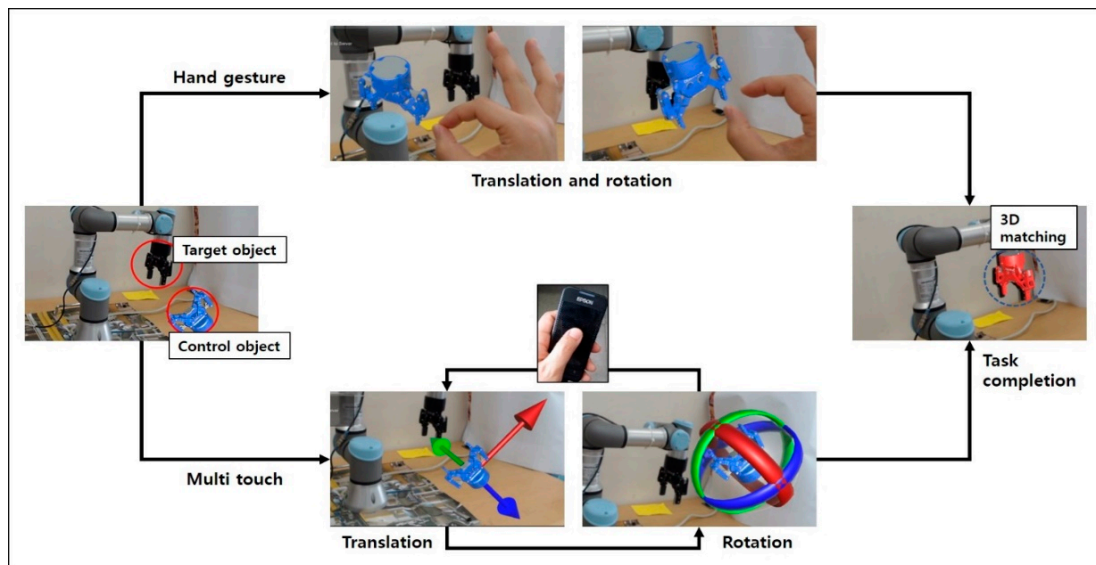


**Figure 13.** Dragging and dropping task for manipulating the robot with a slide bar for changing joint angles and moving the robot along an axis.

#### 4.3.3. 3D Matching Between a Virtual Object and its Corresponding Physical Object

The 3D matching task requires sophisticated user interaction, such as simultaneous 3D translation and rotation, to perform an exact matching between control and target objects in the wearable AR space. In this task, the control object is virtual, and the target object is real. In particular, to quantitatively evaluate the matching between them, an invisible virtual model of the physical target object is created and synchronized at the same position of the physical object. Therefore, the 3D matching task can be performed between a virtual control object and the invisible virtual model of a physical target object, as shown in Figure 14. As one of the fundamental and complicated manipulation tasks in wearable AR, this task was conducted to evaluate both visual contexts and interaction gestures. Note that 3D matching requires both 3D translation and 3D rotation so that it can cover selection and dragging and dropping tasks. However, it is essential to evaluate 2D-based tasks in wearable AR as 2D-based tasks are conducted most frequently.

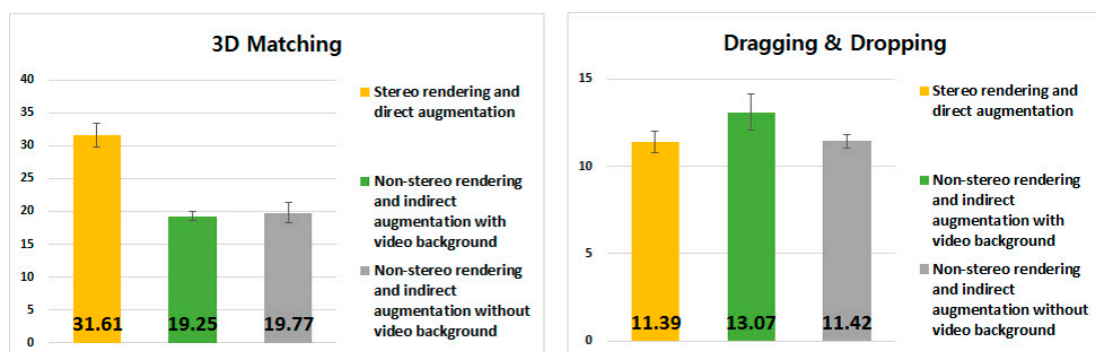
The subject has to recognize the position and orientation of the physical object and has to find the virtual object. After understanding the configuration, the subject selects a virtual control object with the hand gesture or touch gesture, then translates and rotates it, and finally matches it with the target object. In the hand gesture interaction, the control object is selected by conducting a pinch gesture (Figure 6b) with the index finger and thumb. Whenever the subject's hand is moved and rotated, the control object is also moved and rotated. In the case of touch interaction, the subject performs a translation and rotation operation separately since the touch interaction cannot support both operations simultaneously. For this reason, the translation mode and the rotation mode are provided. The task is completed when the control and target objects are matched geometrically.



**Figure 14.** 3D (three-dimensional) matching task between a virtual control object and a physical target object; the gripper of the UR3 robot is used for 3D matching.

#### 4.4. Experimental Results

First, for the analysis of task performance depending on visual contexts, an ANOVA was performed for three visual augmentations. Note that most of the data satisfied ANOVA assumptions. However, some of them did not satisfy ANOVA assumptions (the assumption of homogeneity of variances) due to the limited number of experimental data. For this reason, we used Welch’s ANOVA tests rather than general ANOVA tests through the study [46]. As shown in Figure 15, there was a significant difference ( $p < 0.05$ ) for the 3D matching task, which is complicated and challenging. On the other hand, there was no statistical difference in the dragging and dropping task, which is rather a simple task. The non-stereo rendering and indirect augmentation with video background outperformed the stereo rendering and direct augmentation. The main reason for these results is that the 3D matching task requires much attention and sophisticated manipulation of the control object, so the FOV played an essential role in task performance. On the other hand, the FOV did not influence the performance of the dragging and dropping task very significantly.



**Figure 15.** Task completion time (s) of user interaction for visual contexts.

The qualitative analysis was conducted based on the validated ten questions from the SUS questionnaire [45]. Summarized ten questions in the SUS questionnaire include (1) frequent system use, (2) unnecessarily complex, (3) easy to use, (4) need technical support, (5) functions well integrated, (6) too much inconsistency, (7) easy to learn, (8) very cumbersome, (9) confident using the system, and (10) need to learn many things beforehand. However, original questions without any modification

were given to the subject and used for the questionnaire evaluation. Readers are referred to see the questions in Reference [45].

As shown in Figure 16, there were statistical differences in all questions except Q2 and Q4. Overall, the non-stereo rendering and indirect augmentation with video background showed better qualitative evaluation than the stereo rendering and direct augmentation and the non-stereo rendering and indirect augmentation without video background. The non-stereo rendering and indirect augmentation without video background showed the worst usability evaluation.

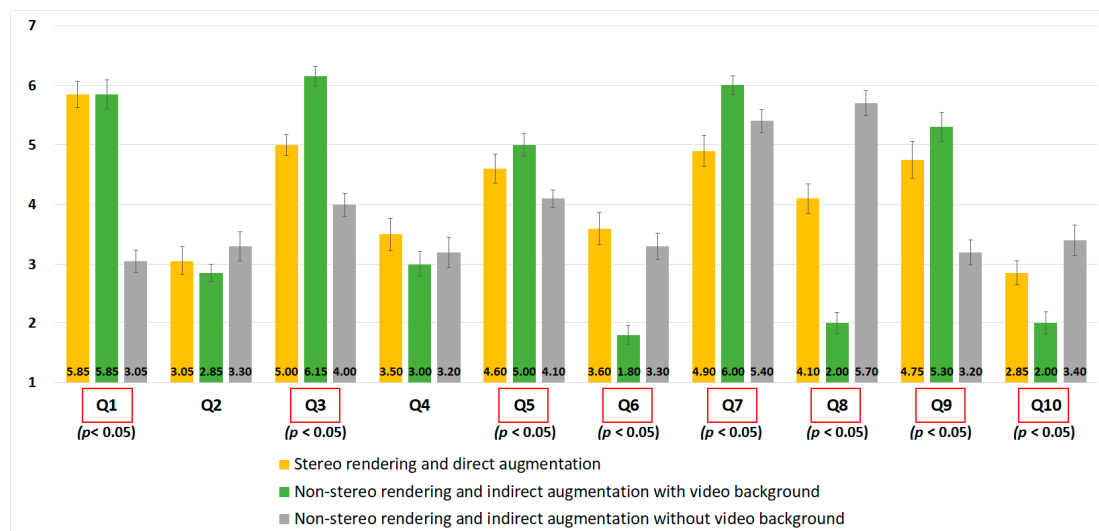


Figure 16. Qualitative result for visual contexts.

For the post hoc analysis concerning the visual context, in Q1 and Q9, there was no difference between the stereo rendering and direct augmentation and the non-stereo rendering and indirect augmentation, but there was a statistical difference between the non-stereo rendering and indirect augmentation without video background and the others concerning the visual context ( $p < 0.05$ ). In particular, in Q3 and Q8, there was a significant difference between the three methods. The non-stereo rendering and indirect augmentation showed better easy-to-use and less cumbersome-to-use ratings. In Q6, Q7, and Q10, there was no difference between the non-stereo rendering and indirect augmentation and the stereo rendering and direct augmentation. We found that, although the stereo rendering and direct augmentation could match the virtual object with the real object and provide immersive visualization, its task performance was worse than the non-stereo rendering and indirect augmentation due to the narrower FOV. We found that the amount of information provided to the subject was a more influential factor than FOV in currently available AR smart glasses, although some mismatch problems did exist. Nevertheless, all the subjects agreed that visual augmentation using AR smart glasses is a promising tool for task assistance.

For the quantitative analysis, according to interaction gestures in wearable AR, a *t*-test was performed for task completion time (comparison between two methods). There was a statistical difference ( $p < 0.05$ ) between interaction gestures in the 3D matching task, but there was no statistical difference in the selection task, as shown in Figure 17. In particular, in the case of hand gesture interaction, the 3D matching was six times faster than the multi-touch interaction using the touchpad.

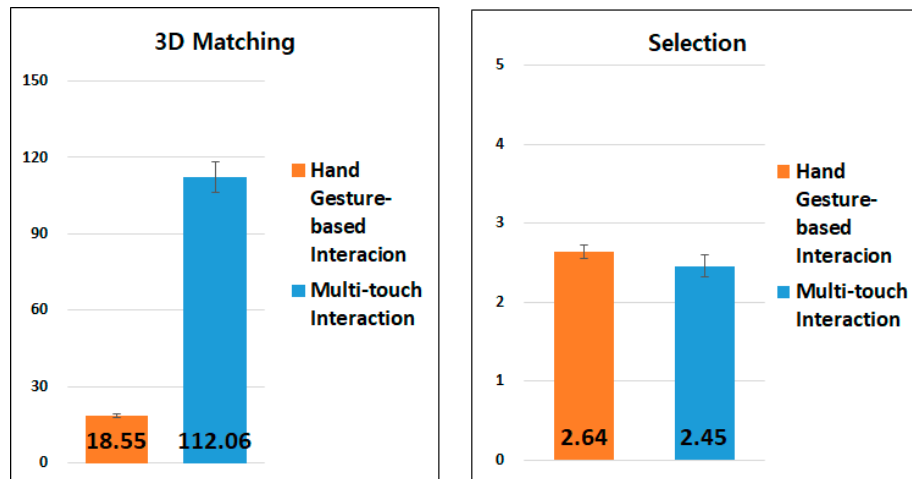


Figure 17. Task completion time (s) for conducting tasks concerning gesture interactions.

For the qualitative analysis of the interaction method, the hand gesture interaction method showed better results than the multi-touch interaction method in all questions except Q4 (need-technical-support), as shown in Figure 18. According to interviews, most participants mentioned that they could easily learn the method and thus could perform tasks more easily and effectively.

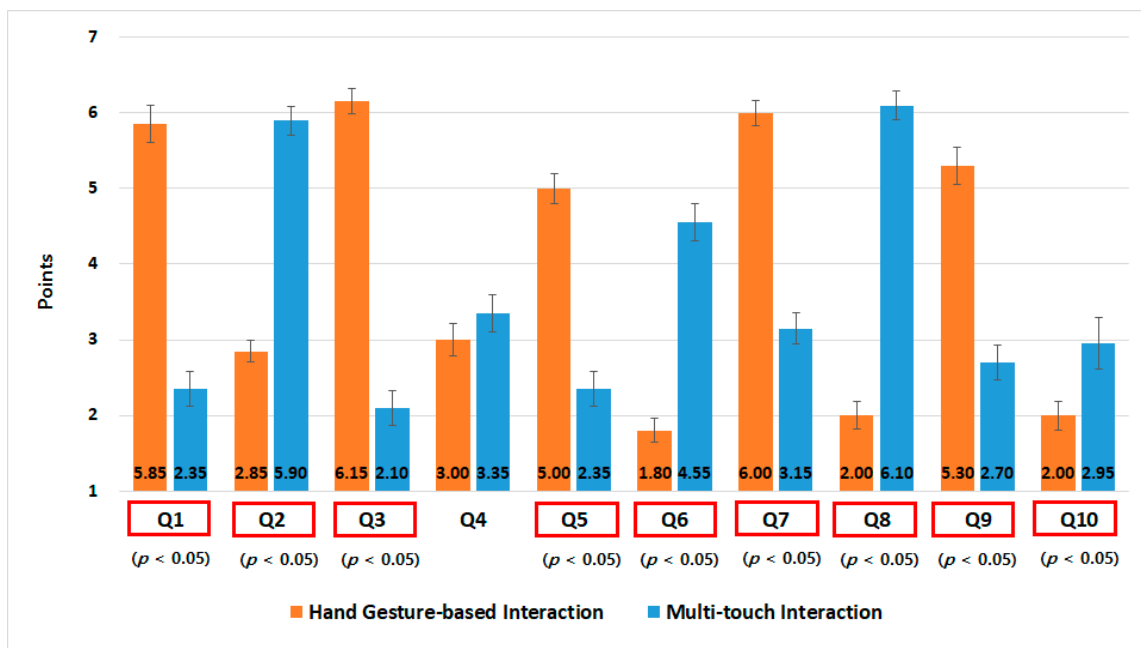


Figure 18. Qualitative results for gesture interaction.

We performed quantitative and qualitative analyses concerning visual contexts and gesture interactions. In particular, the most complicated and difficult task among the three tasks is 3D matching. 3D matching was conducted to evaluate both visual contexts and gesture interactions. For currently available AR smart glasses, such as Epson-Moverio, the non-stereo rendering and indirect augmentation method with video background using the hand gesture outperformed the other methods. One of the main reasons is that the AR smart glasses cannot provide wider FOV enough to provide with all the visual information using the stereo rendering and direct augmentation. We found that it would be more valuable to analyze the relation between gesture interactions and visual contexts. However, we have found that there are inherent limitations to implement them using Epson Moverio



effectively. For this reason, further study is still needed to evaluate the relation between them using different AR smart glasses.

## 5. Comparative Evaluation of Different FOVs

We performed an additional comparative evaluation for the 3D matching task using AR smart glasses with different FOVs. This task was similar to the task performed in Section 4. The difference is that, in Section 4, FOV was defined based on visual contexts on the same see-through display. On the other hand, in Section 5, different AR smart glasses were used. HoloLens provides wider FOV based on the hardware specification.

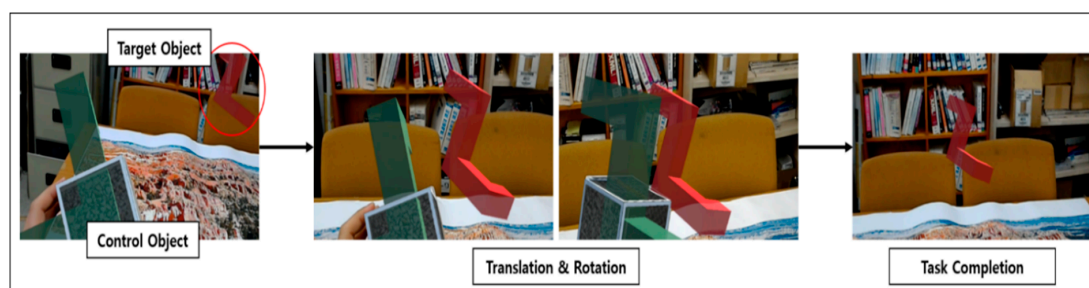
Note that, in this experiment, the control object is physical, and the target object is virtual. In other words, the user holds a physical control object instead of a virtual control object and tries to move and rotate it for 3D matching with a virtual target object. One of the main reasons to perform this experiment is that sometimes workers must hold physical tools and parts to complete assembly and inspection operations in manufacturing environments. Visual guidance and virtual objects are useful to conduct the required task effectively. In this comparative evaluation, the indirect augmentation with and without video background was excluded since it showed worse performance than the others in the previous assessment. As mentioned in Section 4.4, this method showed the worst quantitative and qualitative evaluations. In particular, the task performance concerning the different sizes of FOV was measured for the stereo rendering and direct augmentation.

### 5.1. Experimental Setup and Design

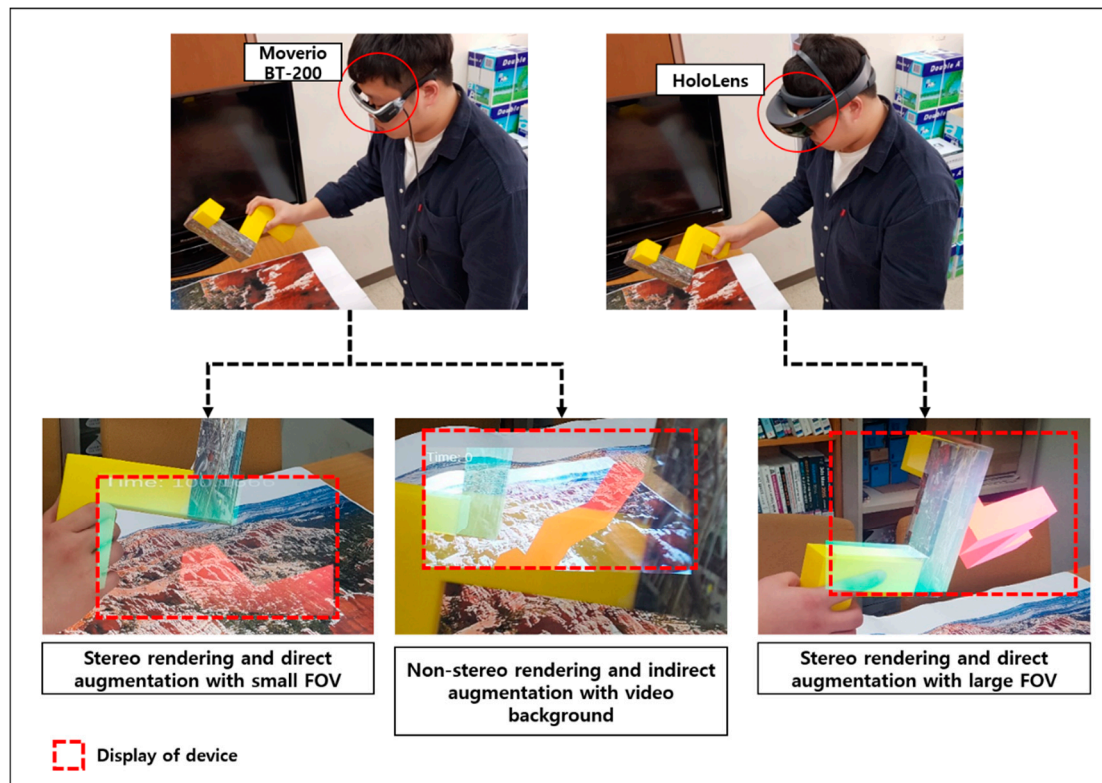
This evaluation was conducted in the same environment as the first experiment. In addition to Epson Moverio BT-200, an MS HoloLens with a larger FOV was used to evaluate the stereo rendering and direct augmentation.

Another fifteen college students (twelve male and three female students) with an average age of 26.3 (SD = 1.5) participated in this user study. No subjects had previously experienced wearable AR. Before the experiment, the participants were also told about the visual interaction methods and related hardware and software according to the visual context. After that, tasks were performed in a random order for the three different methods.

The 3D matching task between the physical control and virtual target objects was conducted to evaluate the task performance, as shown in Figure 19. This experiment was designed to analyze how the user could recognize the position and rotation of the target object (virtual object) and how to accurately match the target object using the control object (physical object). The control and target objects have unique shapes, as shown in Figures 19 and 20. In particular, the shape of the physical target object used in this study was the same as that used in a previous research work [47], which studied the user perception of matching two objects.



**Figure 19.** Additional 3D matching task between the physical control object and the virtual target object.

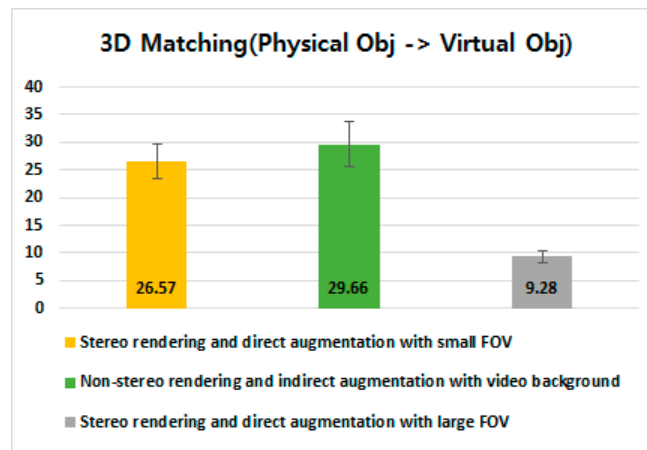


**Figure 20.** The difference of FOV depending on visual contexts and different AR smart glasses (Epson Moverio BT-200 vs. HoloLens).

When the user holds the physical control object and looks at it through the optical see-through display of the AR smart glasses, the virtual model is augmented on the physical control object, and the virtual target object is also visually augmented in the wearable AR space. For each experiment, the target object is located in an arbitrary location with a different orientation. After the user recognizes the position and orientation of the target object through the smart glasses, they translate and rotate the physical object to match with the target object. When the control and target objects are matched, the task is completed. The condition for the successful matching is based on the translation distance and orientation distance between two objects. When both differences are within the given tolerance, the task is completed.

## 5.2. Experimental Results

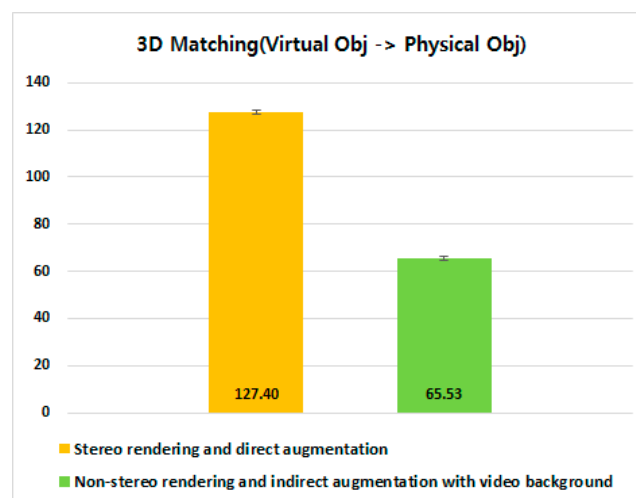
A comparative evaluation of visual interaction according to the visual context using different smart glasses was performed. A Welch's ANOVA was performed on the completion time for task execution. The analysis verified that there was a significant difference ( $p < 0.05$ ) in task completion according to the visual context and FOV. According to the post hoc analysis, there was no statistical difference between the stereo rendering and direct augmentation method with a small FOV and the non-stereo rendering and indirect augmentation with video background method, but there was a considerable difference in the stereo rendering and direct augmentation method with a larger FOV ( $p < 0.05$ ), as shown in Figure 21.



**Figure 21.** Task completion time (s) for 3D matching from a physical object to a virtual object depending on visual contexts and FOV.

### 5.3. 3D Matching Between Virtual Control and Physical Target Objects (Revisited)

The completion time using the stereo rendering and direct augmentation in the first experiment (Figure 15) took longer than the non-stereo rendering and indirect augmentation with a small FOV. However, there was no significant difference between the two augmentations in the second evaluation (Figure 21). We assumed that the different result was caused by whether the control object was real or virtual. To confirm this assumption, an additional experiment was performed by ten participants (male: six, female: four, mean age: 29.3 years old, SD: 4.9) [44]. The additional experiment was performed with 3D matching in the same environment as the second experiment. However, unlike the second experiment, a virtual object was used as a control object to match the physical object like in the first experiment. A quantitative analysis of completion time was performed. We found that there was a statistically significant difference between the two augmentations, as shown in Figure 22. The stereo rendering and direct augmentation method had approximately double the completion time of the non-stereo rendering and indirect augmentation method ( $p < 0.05$ ). Note that we obtained similar results for task performance with the virtual control object for 3D matching in Section 4. The conclusion based on the results of the two experiments is as follows: Unlike the virtual object, when the physical control object is used, it is not necessary for the user to continuously look at the 3D matching process as the user can intuitively manipulate the physical control object. On the other hand, the user must look at the virtual control object continuously to perform the 3D matching task.



**Figure 22.** Task completion time (s) for 3D matching.

## 6. Real Use Case Analysis for Industrial Workers

The experiments in Sections 4 and 5 were performed in a controlled laboratory environment. However, because there might be a difference between the controlled environment and the actual manufacturing environment, an additional qualitative evaluation was performed on industrial workers of a small and medium-sized company that assembles air conditioners and heaters, as shown in Figure 23.

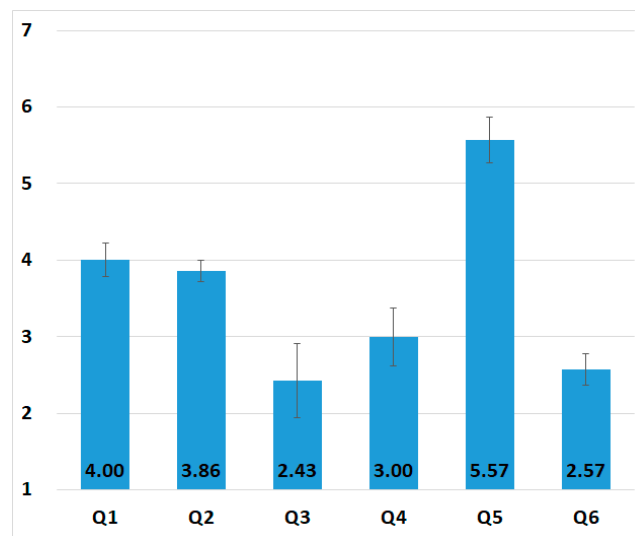


**Figure 23.** Industrial workers were experiencing task assistance through AR smart glasses.

For this real use case analysis, seven industrial workers (average age: 29.7 years old, SD: 3.59) participated in the experiment, and none of them had any prior knowledge of smart glasses. First, we explained the purpose of the experiment to the workers and presented the concept, methods, and functionality of devices, such as smart glasses, visual context, and interaction methods. Then, the workers experienced the task assistance manual required for the actual assembly through smart glasses. Besides, we showed related AR manufacturing services through videos and animations.

After the experiment, the workers were asked to fill out the questionnaire. We analyzed the cognitive load in performing manufacturing tasks since the real industrial field is different from a controlled environment, and wearable AR should not provide additional cognitive burden. We conducted another qualitative evaluation based on the validated questions from NASA task load index (NASA-TLX) (1: very low, 7: very high) [48]. Questions in the NASA-TLX questionnaire include (1) required-mental and perceptual activity, (2) required-physical activity, (3) time pressure felt, (4) hard to work, (5) successful in accomplishing goals, and (6) stressed. We also used these questions without any modifications. Details are shown in Reference [48]. According to the questionnaire analysis (Figure 24), subjects mentioned that they could perform given tasks with less cognitive load with the help of wearable AR. They were satisfied with the visualization and interaction of manufacturing information using AR smart glasses. Some of them commented that wearable AR was very effective when used for training or task assistance with beginners.

However, although AR smart glasses were useful for task assistance, the workers also mentioned that currently available AR smart glasses were heavy to wear, and the screen was not visible in bright places. Also, for the visual context, they preferred the stereo rendering and direct augmentation rather than the non-stereo rendering and indirect augmentation as there was no mismatch between the physical and virtual objects. However, it was suggested that improvement should be made to the fact that much information is not visible at any given time. For the interaction method, they commented that it would be easier to perform the touch interaction in the process of performing simple tasks. However, they mentioned that using two hands was effective in the actual manufacturing environment.



**Figure 24.** Evaluation result based on NASA-TLX of the real use case.

## 7. Discussion on User Studies and Suggestions for User-Centric Task Assistance

The experimental results have proved that the task performance of the user interaction is influenced by both the visual context and the gesture interaction.

### 7.1. Discussion on User Studies

#### 7.1.1. Evaluation for the Visual Context

For the 3D matching task concerning the visual context in the first user study of Section 4, the stereo rendering and direct augmentation method took approximately twice as long to complete the task compared to the non-stereo rendering and indirect augmentation. The reason for this is that the non-stereo rendering and indirect augmentation can provide more information than the stereo rendering and direct augmentation due to the broader FOV through the camera than through the optical display. Due to this limitation, in the stereo rendering and direct augmentation, the user had to move their head frequently to check the location of the target object to select, translate, and rotate the virtual control object.

In the second user study in Section 5, the stereo rendering and direct augmentation with a wider FOV (but still similar or narrower than a camera) showed the best task performance. In this experimental evaluation, the stereo rendering and direct augmentation with a larger FOV made it possible for the user to recognize the position and orientation of the target object easily, and the matching task with the control object was more effectively performed. In particular, if the FOV of the stereo rendering and direct augmentation and that of the non-stereo rendering and indirect augmentation were similar, the stereo rendering and direct augmentation was more effective than the non-stereo rendering and indirect augmentation due to the visual matching between virtual and physical artifacts. On the other hand, the non-stereo rendering and indirect augmentation allowed the physical objects to be visually overlapped only through the display of the smart glasses. Thus, the video image was mismatched with physical objects, confusing the user while they performed their tasks.

Depending on the role of the control and target objects for the 3D matching task in the first and second experiments, there was a difference in task performance. The reason for this result is related to whether the control object with which the user interacts is virtual or real. For the 3D matching in the first user study, the control object was virtual. Thus, when the user moved and rotated the control object, the control object always had to be in the user's field of view. However, as the control object in the second evaluation was real, it was not necessary to place the control object in the user's field

of view in the AR smart glasses. That is, when the user recognized the location and orientation of the target object while holding the control object, the task could be conducted without looking at the control object.

According to the qualitative analysis and the interview for the visual context-based methods, the stereo rendering and direct augmentation method provides a promising direction for wearable AR as it can support almost exact matching between the virtual and physical objects in wearable AR. However, some participants mentioned that the stereo rendering and direct augmentation also had a limitation in providing enough information due to the narrow FOV, which prevented them from performing the given tasks naturally. In the non-stereo rendering and indirect augmentation with the video background, physical objects were duplicated through the display of the smart glasses. Interestingly, some users performed tasks just by looking only at the display of the smart glasses in the video-based augmentation. On the other hand, the non-stereo rendering and indirect augmentation without background was not suitable for the task assistance as there was much difference between the virtual object and the real environment. We also found similar results from industrial workers in a real industrial environment. Therefore, the analysis and evaluation results will be effectively utilized in designing and implementing manufacturing services using AR smart glasses.

#### 7.1.2. Evaluation for the Gesture Interaction

For the 3D virtual object matching task concerning different gesture interactions, the hand gesture interaction demonstrated faster task completion time than the multi-touch interaction. The faster task completion is due to the concurrency control of the interaction (selection, translation, rotation, etc.) and intuitiveness of using a real hand to perform the task. Since the hand gesture-based interaction provides the same metaphor as the user performs typical tasks in the real environment, the user can simultaneously perform the translation and rotation without a separate mode change through the selection of the menu using their hand. On the other hand, in the case of the multi-touch interaction, whenever a control object is selected through the touchpad of smart glasses or the transformation needs to be changed, this method requires a lot of interaction with the gizmo. Furthermore, the user must change the mode to select and manipulate the control object whenever a different transformation is required.

Interviews also showed that interaction with hand gestures was more natural and effective. However, it was uncomfortable to wear another sensor for detecting hand gestures. On the other hand, the interaction using the multi-touch interaction made sophisticated tasks difficult because the mode had to be converted sequentially and frequently (for example, in order of selection → object translation → object rotation) to complete a specific interaction. As a result, it was found that the satisfaction level was lower than that of the hand gesture.

There was no statistical difference between different gesture interactions in 2D interactions, such as selection. Interestingly, when performing the 2D-based tasks, some users moved their heads to specific locations and then clicked menus rather than moving the cursor with the touchpad.

### 7.2. Challenges for Supporting User-Centric Task Assistance

Based on the evaluation results, we discussed some considerations and challenges for user-centric task assistances using AR smart glasses.

#### 7.2.1. Visual Layout for Cognitive Information Visualization

It has been shown that due to the narrow FOV of currently available AR smart glasses, the necessary information cannot be sufficiently visualized on the display of smart glasses, and the augmented information may be out of the FOV of the screen. However, for smart glasses to support a wider FOV, their cost must increase.

For this reason, for currently available AR smart glasses, it is necessary to consider a dynamic visual layout for the overlaid information to make it easier for the user to recognize and understand

the information. One of the important considerations for this purpose is to analyze the relationship and priority between visual AR information. Usually, the worker requires a different level of visual information to perform given tasks. Thus, it is essential to provide specific information that is more important for performing the tasks based on the relationship and priority of the AR information. The visual layout, such as the Fisheye view [49,50], calculates the degree of interest of the augmented information and increases the visual focus of the information as the priority is higher. Thus, by providing a more appropriate and important information-focused view by considering worker situations, it is possible to provide more effective task support to workers.

### 7.2.2. Hybrid User Interaction

Visual components augmented in a wearable AR environment are composed of various forms, such as a menu, a widget, and a 3D model. According to the experimental evaluation in this research, the multi-touch interaction is feasible for manipulating 2D virtual objects, e.g., for selection and dragging. On the other hand, the hand gesture interaction is suitable for manipulating 3D models, e.g., for translating and rotating. Therefore, a hybrid user interface can simultaneously support multi-touch interaction to perform various 2D tasks and hand gesture interaction to perform more complex 3D tasks. Besides, as smart glasses evolve, research on interfaces for interaction will be steadily conducted, and more effective interactions will be possible for the hybrid user interface.

### 7.2.3. Enhanced Hand Gesture Interaction

In a wearable AR environment, the user interacts with physical objects, as well as virtual objects, through AR interactions in the physical environment. However, for a virtual control object, the position and orientation of the virtual object are tracked through the camera, and the virtual object should be carefully manipulated with the hand gesture interaction, while the virtual object is augmented within the FOV. For this reason, to manipulate a virtual control object, the user must always look at it. On the other hand, a physical control object does not need to be viewed during the interaction. Therefore, there is a difference between the interaction with the real object and the interaction with the virtual object.

For this reason, we require a new method for supporting more effective manipulation of virtual objects compared with that of physical objects. One way to do this is to track the movement of the user's hand with additional sensors. The translation and rotation of virtual objects can be inferred even if the user does not look at the virtual object through the movement of their hand. For example, after capturing a virtual object through a hand gesture interaction, the moving speed and acceleration information of the user's hand with IMU is tracked to calculate the movement of the virtual object [51]. In this way, it is possible to interact with a virtual object more easily and more like the interaction with a real object.

## 8. Conclusions

As many AR smart glasses are currently available [6], it is essential to make visual augmentation more natural and intuitive in terms of visual contexts and gesture interaction. However, there has been little research work on investigating this issue in depth. In this study, user interactions based on visual contexts and gesture interactions were classified and evaluated using AR smart glasses. We performed quantitative and qualitative analyses, including performance measurement and questionnaire evaluation. The results verified that there was a significant difference in task performance according to visual contexts and gesture interactions as the task complexity increased. They also showed that both FOV and visual registration between virtual and physical artifacts were important.

We have analyzed not only the interaction between virtual objects but also the interaction between physical and virtual objects. Depending on the role of the control and target objects for the 3D matching task, there was a difference in task performance. The reason for this result is related to whether the control object with which the user interacts is virtual or real.

We have also found that both the FOV and the gesture interaction influence task performance, and they can complement each other. However, currently available smart glasses have narrower FOV than is necessary, and the visual matching assistance is not sufficient for use in real-world tasks. Although interviews showed that interaction with hand gestures was more natural and effective, it was uncomfortable to wear another sensor for detecting hand gestures. Nevertheless, we have demonstrated the potential utilization of AR smart glasses. To overcome still existing limitations on the visual context and gesture interaction, it is necessary to provide a dynamic visual layout concerning narrow FOV and to support hybrid user interfaces for more accurate and natural interaction.

Actual manufacturing systems may have different influential factors compared with the controlled laboratory, although a real use case study was performed qualitatively using industrial workers. Therefore, further study is required to analyze the visual context and interaction methods in different real manufacturing environments. For this reason, we plan to apply the proposed approach to a smart factory testbed that produces motors and to evaluate its performance quantitatively and qualitatively. Secondly, it is necessary to evaluate the relation between gesture interactions and visual contexts using different AR smart glasses. Thirdly, we need to improve the accuracy of the hand gesture and visual registration in wearable AR. Fourthly, we need further investigation of different types of AR smart glasses for analyzing visual contexts and gesture interactions concerning hardware specification and price. Finally, it would be valuable for the visualization and interaction methods to be extended to adopt some of the challenges described previously.

**Author Contributions:** Conceptualization, M.K. and J.Y.L.; methodology, M.K., S.H.C., K.-B.P., and J.Y.L.; formal analysis, M.K. and J.Y.L.; investigation, S.H.C., K.-B.P., and J.Y.L.; software, M.K., S.H.C., and K.-B.P.; validation, M.K. and J.Y.L.; supervision, J.Y.L.; writing-review and editing, J.Y.L.

**Funding:** This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (NRF-2016R1D1A1B03934697 and 2019R1I1A3A01059082).

**Acknowledgments:** We thank Hyeon-Seok Kim for the experimental setup.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Zheng, X.S.; Foucault, C.; Da Silva, P.M.; Dasari, S.; Yang, T.; Goose, S. Eye-wearable technology for machine maintenance: Effects of display position and hands-free operation. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 2125–2134.
- Friedrich, W.; Jahn, D.; Schmidt, L. ARVIKA-Augmented reality for development, production and service. In Proceedings of the International Symposium on Mixed and Augmented Reality, Darmstadt, Germany, 1 October 2002; pp. 3–4.
- Westerfield, G.; Mitrovic, A.; Billinghamurst, M. Intelligent augmented reality training for motherboard assembly. *Int. J. Artif. Intell. Educ.* **2015**, *25*, 157–172. [[CrossRef](#)]
- Kim, M.; Lee, J.Y. Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality. *Multimed. Tools Appl.* **2016**, *75*, 16529–16550. [[CrossRef](#)]
- Choi, S.H.; Kim, M.; Lee, J.Y. Situation-dependent remote AR collaborations: Image-based collaboration using a 3D perspective map and live video-based collaboration with a synchronized VR mode. *Comput. Ind.* **2018**, *101*, 51–66. [[CrossRef](#)]
- Syberfeldt, A.; Danielsson, O.; Gustavsson, P. Augmented reality smart glasses in the smart factory: Product evaluation guidelines and review of available products. *IEEE Access* **2017**, *5*, 9118–9130. [[CrossRef](#)]
- Henderson, S.J.; Feiner, S.K. Augmented reality in the psychomotor phase of a procedural task. In Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality, Basel, Switzerland, 26–29 October 2011; pp. 191–200.
- Khuong, B.M.; Kiyokawa, K.; Miller, A.; La Viola, J.J.; Mashita, T.; Takemura, H. The effectiveness of an AR-based context-aware assembly support system in object assembly. In Proceedings of the IEEE Virtual Reality, Minneapolis, MN, USA, 29 March–2 April 2014; pp. 57–62.



9. Zhu, Z.; Branzoi, V.; Wolverton, M.; Murray, G.; Vitovitch, N.; Yarnall, L.; Acharya, G.; Samarasekera, S.; Kumar, R. AR-mentor: Augmented reality based mentoring system. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, Munich, Germany, 10–12 September 2014; pp. 17–22.
10. Moverio BT-200. Available online: [www.epson.com/MoverioBT200](http://www.epson.com/MoverioBT200) (accessed on 5 April 2017).
11. MS HoloLens. Available online: <https://www.microsoft.com/en-us/hololens> (accessed on 20 March 2018).
12. Ergonomics of Human-System Interaction. Available online: <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en> (accessed on 10 October 2018).
13. Robertson, C.; MacIntyre, B. Adapting to registration error in an intent-based augmentation system. In *Virtual and Augmented Reality Applications in Manufacturing*, 1st ed.; Ong, S.K., Nee, A.Y.C., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 147–167.
14. Robertson, C.M.; MacIntyre, B.; Walker, B.N. An evaluation of graphical context as a means for ameliorating the effects of registration error. *IEEE Trans. Vis. Comput. Graph.* **2008**, *15*, 179–192. [[CrossRef](#)] [[PubMed](#)]
15. Longo, F.; Nicoletti, L.; Padovano, A. Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Comput. Ind. Eng.* **2017**, *113*, 144–159. [[CrossRef](#)]
16. Funk, M.; Kritzler, M.; Michahelles, F. HoloCollab: A shared virtual platform for physical assembly training using spatially-aware head-mounted displays. In Proceedings of the Seventh International Conference on the Internet of Things, Linz, Austria, 22–25 October 2017; p. 19.
17. Benko, H.; Ofek, E.; Zheng, F.; Wilson, A.D. Fovear: Combining an optically see-through near-eye display with projector-based spatial augmented reality. In Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology, Charlotte, NC, USA, 11–15 November 2015; pp. 129–135.
18. Nambu, R.; Kimoto, T.; Morita, T.; Yamaguchi, T. Integrating smart glasses with question-answering module in assistant work environment. *Procedia Comput. Sci.* **2016**, *96*, 1772–1781. [[CrossRef](#)]
19. Büttner, S.; Funk, M.; Sand, O.; Röcker, C. Using head-mounted displays and in-situ projection for assistive systems: A comparison. In Proceedings of the 9th ACM international Conference on Pervasive Technologies Related to Assistive Environments, Corfu, Island, Greece, 29 June–1 July 2016; p. 44.
20. Funk, M.; Kosch, T.; Schmidt, A. Interactive worker assistance: Comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 934–939.
21. Ha, T.; Feiner, S.; Woo, W. WeARHand: Head-worn, RGB-D camera-based, bare-hand user interface with visually enhanced depth perception. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, Munich, Germany, 10–12 September 2014; pp. 219–228.
22. Wang, X.; Ong, S.K.; Nee, A.Y.C. Real-virtual components interaction for assembly simulation and planning. *Robot. Comput. Integr. Manuf.* **2016**, *41*, 102–114. [[CrossRef](#)]
23. Wang, X.; Ong, S.K.; Nee, A.Y.C. Multi-modal augmented-reality assembly guidance based on bare-hand interface. *Adv. Eng. Inform.* **2016**, *30*, 406–421. [[CrossRef](#)]
24. Lee, J.Y.; Rhee, G.W.; Seo, D.W. Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment. *Int. J. Adv. Manuf. Technol.* **2010**, *51*, 1069–1082. [[CrossRef](#)]
25. Huang, Z.; Li, W.; Hui, P. Ubii: Towards seamless interaction between digital and physical worlds. In Proceedings of the 23rd ACM international conference on Multimedia, Brisbane, Austria, 26–30 October 2015; pp. 341–350.
26. He, Z.; Yang, X. Hand-based interaction for object manipulation with augmented reality glasses. In Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry, Shenzhen, China, 30 November–2 December 2014; pp. 227–230.
27. Brancati, N.; Caggianese, G.; De Pietro, G.; Frucci, M.; Gallo, L.; Neroni, P. Usability evaluation of a wearable augmented reality system for the enjoyment of the cultural heritage. In Proceedings of the 11th International Conference on Signal-Image Technology and Internet-Based Systems, Bangkok, Thailand, 23–27 November 2015; pp. 768–774.
28. Hsieh, Y.T.; Jylhä, A.; Orso, V.; Gamberini, L.; Jacucci, G. Designing a willing-to-use-in-public hand gestural interaction technique for smart glasses. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 4203–4215.

29. Haque, F.; Nancel, M.; Vogel, D. Myopoint: Pointing and clicking using forearm mounted electromyography and inertial motion sensors. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 3653–3656.
30. Dudley, J.J.; Vertanen, K.; Kristensson, P.O. Fast and Precise Touch-Based Text Entry for Head-Mounted Augmented Reality with Variable Occlusion. *ACM Trans. Comput. Hum. Int.* **2018**, *25*, 30. [CrossRef]
31. Ha, T.; Woo, W. ARWand: Phone-based 3d object manipulation in augmented reality environment. In Proceedings of the International Symposium on Ubiquitous Virtual Reality, Jeju, Korea, 1–4 July 2011; pp. 44–47.
32. Yang, Y.; Shim, J.; Chae, S.; Han, T.D. Interactive Augmented Reality Authoring System using mobile device as input method. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Budapest, Hungary, 9–12 October 2016; pp. 1429–1432.
33. Ahn, S.; Heo, S.; Lee, G. Typing on a smartwatch for smart glasses. In Proceedings of the ACM International Conference on Interactive Surfaces and Spaces, Brighton, UK, 17–20 October 2017; pp. 201–209.
34. Whitmire, E.; Jain, M.; Jain, D.; Nelson, G.; Karkar, R.; Patel, S.; Goel, M. Digitouch: Reconfigurable thumb-to-finger input and text entry on head-mounted displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 113:1–113:21. [CrossRef]
35. Wang, C.Y.; Chu, W.C.; Chiu, P.T.; Hsiu, M.C.; Chiang, Y.H.; Chen, M.Y. PalmType: Using palms as keyboards for smart glasses. In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, Copenhagen, Denmark, 24–27 August 2015; pp. 153–160.
36. Lissermann, R.; Huber, J.; Hadjakos, A.; Nanayakkara, S.; Mühlhäuser, M. EarPut: Augmenting ear-worn devices for ear-based interaction. In Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design, Sydney, New South Wales, Australia, 2–5 December 2014; pp. 300–307.
37. Raynal, M.; Gauffre, G.; Bach, C.; Schmitt, B.; Dubois, E. Tactile camera vs. tangible camera: Taking advantage of small physical artefacts to navigate into large data collection. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, Reykjavik, Iceland, 16–20 October 2010; pp. 373–382.
38. Terrenghi, L.; Kirk, D.; Sellen, A.; Izadi, S. Affordances for manipulation of physical versus digital media on interactive surfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, QC, Canada, 22–27 April 2006; pp. 1157–1166.
39. Unity3D. Available online: <http://unity3d.com> (accessed on 17 December 2017).
40. Vuforia SDK. Available online: <https://www.vuforia.com/> (accessed on 18 December 2017).
41. Leap Motion. Available online: <https://www.leapmotion.com/> (accessed on 27 December 2017).
42. Moser, K.R.; Swan, J.E. Evaluation of user-centric optical see-through head-mounted display calibration using a leap motion controller. In Proceedings of the IEEE Symposium on 3D User Interfaces, Greenville, SC, USA, 19–20 March 2016; pp. 159–167.
43. Umeyama, S. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *4*, 376–380. [CrossRef]
44. Nielsen, J. Why You Only Need to Test with 5 Users? 2000. Available online: <http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> (accessed on 3 December 2018).
45. Bangor, A.; Kortum, P.; Miller, J. Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usability Stud.* **2009**, *4*, 114–123.
46. Welch's ANOVA Test. Available online: <https://www.real-statistics.com/one-way-analysis-of-variance-anova/welchs-procedure/> (accessed on 28 February 2019).
47. Shepard, R.N.; Metzler, J. Mental rotation of three-dimensional objects. *Science* **1971**, *171*, 701–703. [CrossRef] [PubMed]
48. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv. Psychol.* **1988**, *52*, 139–183.
49. Furnas, G.W. Generalized fisheye views. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 13–17 April 1986; pp. 16–23.

50. Belkacem, I.; Pecci, I.; Martin, B. Smart Glasses: A semantic fisheye view on tiled user interfaces. In Proceedings of the Federated Conference on Computer Science and Information Systems, Gdańsk, Poland, 11–14 September 2016; pp. 1405–1408.
51. Lee, L.H.; Hui, P. Interaction methods for smart glasses: A survey. *IEEE Access* **2018**, *6*, 28712–28732. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).