



A novel adaptive visualization method based on user intention in AR manual assembly

Yuxiang Yan¹ · Xiaoliang Bai¹ · Weiping He¹ · Shuxia Wang¹ · Xiangyu Zhang¹ · Peng Wang² · Liwei Liu¹ · Qing Yu¹

Received: 1 June 2023 / Accepted: 20 October 2023 / Published online: 10 November 2023
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

Abstract

AR assembly instruction is a method that helps users with manual assembly by overlaying virtual information on the real environment. However, users have to manually change the information display form, which fails to adapt to their varying information needs in different stages of the assembly task. This study proposes a user-centered adaptive visualization method for AR manual assembly that can provide the necessary information according to the user's intention in the assembly process. This paper develops a system (UIAVS) through an information hierarchy mechanism and a user intention recognition method based on the user's eye gaze, the spatial location, and current assembly task, which can adaptively adjust the visibility of information, how much information is displayed, and the change of visualization form. UIAVS was tested for the first time in a small engine assembly task. Twenty-four participants were randomly assigned to different task familiarity (novice and expert) and different information visualization response methods for assembly task experiments. The results show that UIAVS has better task performance and cognitive performance as well as user experience than traditional AR assembly instruction method, especially for experts. The research results have certain guiding significance for the design of AR assembly instructions, which extends the application of AR technology in practical complex assembly tasks.

Keywords User intention · Eye gaze · Spatial location · Adaptive · Visualization · Augmented reality

1 Introduction

Augmented reality (AR) is a technology that can integrate the physical world with the virtual world and change the way humans interact with digital data [1]. Compared with the traditional assembly process, augmented reality can display the assembly guidance information fused with the actual work scene in real time, such as step-by-step instructions, 3D models, or other relevant information [2]. The visual information of augmented reality is intuitive and easy to understand [3]. In addition, people can naturally interact with virtual information, which effectively improves the

efficiency of manual assembly [4]. Traditional paper-based assembly processes, which mainly use two-dimensional text and pictures to convey process information, are still widely employed on assembly sites. However, this method of expressing information is flat, static, and limited, and it challenges the users' cognitive abilities when they encounter complex assembly scenarios [5, 6]. Model-based definition (MBD) is a method of defining product assemblies using 3D models (e.g., solid models), product and manufacturing information (PMI), and relevant metadata [7]. Unlike traditional methods that rely on 2D engineering drawings to provide assembly information, MBD can express assembly intent and requirements more clearly and present them on an electronic screen in a way that enhances the user's understanding of the assembly process through views, animations, annotations, and other visual forms [8]. However, these methods of expressing assembly process information cannot integrate the information with the assembly site in a virtual and real way. The expression and interaction of the information do not consider the context and cannot change automatically with the environment. AR assembly instructions are a guidance method that uses augmented reality

✉ Xiaoliang Bai
bxl_nwpu@163.com

✉ Weiping He
weiping@nwpu.edu.cn

¹ Cyber-Physical Interaction Lab, Northwestern Polytechnical University, Xi'an 710072, China

² School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

technology to assist manual assembly. It shows the user the assembly steps, tools, and components by overlaying information such as virtual 3D models, images, animations, and text in a realistic environment [9]. AR assembly instructions differ from traditional paper-based processes and MBD processes in that they are a three-dimensional, dynamic, and unlimited way of expressing information [10–12]. They can automatically provide and update information based on the user's interaction feedback and environmental changes, and change the content of the information in real time [13]. As a result, the information display is more intuitive and easier to understand, and the user does not need to search for information elsewhere [14–16]. Previous studies [6, 17] have shown that AR assembly instructions have advantages over other methods of displaying information in terms of improving assembly efficiency and reducing user cognitive load.

Most previous studies [18–20] aimed to investigate the feasibility of applying AR to manual assembly assistance and demonstrate its effectiveness. However, these studies varied greatly in terms of user interface settings and visual information display [9]. Therefore, how the display of visual information affects the effectiveness of AR applications remains to be explored. Displaying all the visual information will negatively affect users, because they will struggle to cope with the complex multitask information in the virtual and real world [21, 22]. In complex tasks, displaying excessive AR assembly information may make it difficult for users to find the required information, thus prolonging the time to complete the assembly task [23]. In contrast, in simple tasks, users may have low dependence on AR instructions, or even no noticeable effect [6]. Radkowski et al. [23] suggested that the complexity of visual information should match the complexity of assembly steps. With existing methods [24, 25], users need to manually adjust the information they see: assembly step information, assembly parameters, and the display of changes in the 3D model. Switching to different stages of the task (for example, when the user moves from a distance to the final assembly position of the part) requires readjusting the visualization type and level of detail (LOD) of the part information. Users have to manually adjust multiple interactive operations to find the desired information or visual form. This process is cumbersome, especially when users do not even know what interaction to do to find the information they want, it is ineffective and a waste of time.

In fact, users with different levels of expertise have different information needs at different stages of the assembly task [26]. Experts need less information while novices need more information to complete the task successfully. Therefore, a flexible visualization method is needed at the assembly site to accommodate their varying needs in different assembly steps. Moreover, a lack of adaptive experience can adversely affect the overall experience. This paper argues that it is necessary to adaptively adjust how much information is

displayed and in what visual form according to the user's intention. This study aims to propose a user-centered adaptive visualization method that can offer the necessary information based on the user's intention in the assembly process, thus avoiding information overload, reducing cognitive load, and improving assembly efficiency.

Current research on the adaptive display of virtual information in AR assembly focuses on the method of automatically adjusting the display interface to adapt to the current environment according to context awareness [25], or automatically allocating resources and dynamically displaying different contents based on user preferences [27]. Most of these methods only display the amount of information according to the cognitive load, users' preferences, and the difficulty of the task, but they do not highlight the information users need. However, none of these works take user intention into account, that is, they do not adjust the visualization forms of virtual information adaptively and jointly based on user behavior (such as spatial location and eye gaze) in manual assembly. User's spatial location is an important indicator of the area that user is interested in [28]. In addition, eye gaze is directly related to the attentional process and implicitly tells us what we are interested in [29]. The user's intention and status can be dynamically tracked by sensing the gaze point and gaze time through sensors [30]. This paper argues that these aspects play an equally important role in the adaptive display of virtual information, especially in the field of AR manual assembly.

To our best knowledge, this paper first proposes a method to adaptively adjust the display of virtual information according to the user's intention in AR manual assembly. This method can recognize user intention according to the user's eye gaze, current task, and user's spatial location, and adaptively adjust the way of virtual information display to highlight the information required by the user. This paper combines user's intention and visual form with the content of assembly process based on the process designer decision to automatically control when virtual information is displayed, how much information should be displayed, and what visualization form should be used.

Concretely, the system design rules are:

- (1) The design methods of assembly process contents with different LODs and visualization forms shall be determined by the process designer.
- (2) The adaptive display rules of virtual information are designed from the two dimensions of time (user's eye gaze time) and space (user's spatial location), so that different LODs and visual forms can be displayed adaptively according to the user's intention. At runtime, the system takes the user's eye gaze information, the user's spatial location, and the current assembly task into account. By quantifying these factors, the system

recognizes the user's intention and connects them with the visualization information to adaptively control the display form and amount of virtual information.

In summary, this paper contributes an adaptive visual enhancement method of virtual information based on user intention in AR manual assembly. The method can adaptively change the display form of virtual information according to the user's intention, and display the information required by the user to reduce people's cognitive burden. The main contributions of this paper include:

- Proposing an adaptive visualization method of virtual information based on user intention in AR manual assembly.
- Designing a logic rule-based user intention recognition method that can infer user's intention based on user's spatial location, user gaze, and current assembly task in AR manual assembly.
- Reporting on a formal user study investigating the benefits of User-Intention Adaptive Visualization System in AR manual assembly.
- Providing a novel design interface that enriches the design of AR assembly instructions.

The rest of the paper reviews related work, and then describes the methodology, mainly focusing on the design of the system framework and the recognition of user intention as well as the adaptive visualization of virtual information in AR assembly. Next, it reports on a user study with the platform, and discusses the results found. Finally, it provides some conclusions and directions for future work.

2 Related work

This research investigated the method of adaptive display of virtual information based on user intention in AR assembly. Here, this paper reviewed the visualization methods of virtual information in AR assembly and the relevant research methods of adaptive augmented reality, then reviewed the related work on the perceived form of user intention in AR, and finally compared our approach to these prior works.

2.1 Virtual information visualization in AR assembly

Compared with the traditional instructions, the information display method in AR assembly is more intuitive. It uses two-dimensional graphics, text, images, three-dimensional models, animation, and other information to describe the state, assembly relationship, and assembly process of parts in physical tasks [10]. Operators can concentrate on

completing assembly tasks quickly and with high quality [31, 32]. In the AR-based cable routing system, the cable routing direction and text annotation information were superimposed on the vision of the routing operator to help them complete the routing task [33], which was one of the most famous AR application cases in the assembly field. VTT technology research center of Finland had developed an assembly prototype system based on augmented reality, which used virtual parts, virtual assembly tools, and assembly prompts to guide users in actual assembly operations [20]. At present, most of these researches in AR assembly focus on the display of geometric assembly relations, and rarely involve the description and display of semantic relations of part assembly.

Semantic information plays an important role in assembly guidance, which can describe assembly sequence, assembly constraints, assembly tools, and so on. Aiming at the problem that the geometric model cannot fully express the part assembly relationship, Xia et al. [31] proposed an analysis and construction method of part assembly semantic information model based on part object model, visually modeled the assembly relationship between parts, and accurately superimposed these information models on the real assembly environment by using augmented reality and virtual space modeling technology. Wang et al. [34] converted assembly semantic information into more understandable information based on visual content and visual form to improve the human-computer cooperation experience and improve the user's cognitive efficiency of the operation intention of the physical task.

Most of the above works focus on the representation of visual information. For the variable display of visual information, users mostly change the visual form through explicit interaction [35]. It gives users great freedom to get the information they want. However, too many interactive operation interfaces and cumbersome operations will bring information disorder and operational confusion to users, especially in the AR assembly environment with complex background information. In terms of the response to visual information in AR assembly, Radkowski et al. [23] proposed that the visual features of specific assembly operations must correspond to their relative difficulty level, and the ultimate goal was to associate different types of visual features with different levels of task complexity. Lindlbauer et al. [25] also proposed a method to intelligently adjust which applications were displayed, how much information was displayed, and where they were placed based on users' current cognitive load and knowledge about their tasks and environment.

However, to our best knowledge, few works associate the display of virtual information with user intention in AR assembly. In fact, when users assemble in different positions, different operation steps, and different parts, they pay attention to different key information. Generally speaking, users

pay more attention to the information most relevant to the current assembly task. This study takes into account user intentions based on eye gaze, tasks, and user space position, and adaptively adjusts the display form of virtual information to provide users with the most needed information.

2.2 Adaptive augmented reality

In the assembly process of a AR physical task, the superposition of a large amount of virtual information in the real environment, especially there may be many objects in the assembly site of large parts, will cause information redundancy and make it difficult for users to pay attention to the key information needed. In order to avoid information overload, the AR system can be designed as an explicit interactive system or an implicit interactive system [36]. Explicit interaction systems can use gestures, finger pointing [37], or other input devices to provide the precise information we need. The implicit interactive system can provide us with a method of adaptive display of information, and avoid manual operation by designing an active interactive way. The display method of implicit interactive information is mainly through context awareness [29], and context awareness factors include environmental knowledge (object position/shape, etc.), or approximate relevant objectives and rules [38, 39]. In order to follow this approach, various augmented reality (AR) systems have been developed [40, 41]. DiVerdi et al. [42] proposed a method based on the combination of level of detail (LOD) geometry and adaptive user interface. LOD interfaces allowed applications to adjust the size of the interface widget with respect to distance from the camera to make good use of the screen space in the 3D environment. They proposed to use different LODs of content presentations to improve the adaptability of the augmented reality interface. The concept of LOD is also used in this research. In order to overcome the confusion caused by overlapping and complex information when displaying virtual information, Julier et al. [38] proposed a method to perform information filtering by using region-based and users' location and task. Tatzgern et al. [43] presented an adaptive information density display method for AR to reduce clutter when the amount of data is too large, so that users can focus on finding the relevant items. Ghouaiel et al. [44] proposed to adjust the virtual content according to the changes of ambient light, target distance, and noise. In the field of AR assembly, Geng et al. [27] developed a system, which can adaptively allocate resources and dynamically display different contents according to the preferences of different operators. Lindlbauer D et al. [25] proposed a method to automatically adjust the augmented reality interface based on the user's current cognitive load and knowledge about their task and environment. This method uses a mix of rule-based decision-making and combinatorial optimization to determine when, where,

and how to display virtual elements. For tasks with low cognitive load, the system displays more elements and more details. With the increase of cognitive load, the minimum user interface is adjusted to display fewer elements on tasks with high cognitive load. Although the above method of adaptive display of information based on context awareness can adjust the display form of information, one limitation of this implicit method is approximate correlation, which is not necessarily related to the real intention [29]. These methods cannot adaptively display the visual enhancement information that users are interested in according to the user's intention. The ideal interface design is to support two ways, that is, to display approximate relevant information in an implicit way, and to provide an explicit way to interact with it to take into account various factors. Pfeuffer et al. [29] believe that gaze interaction is an ideal way to realize this design. They have created a gaze interaction design space to continuously expand more content according to users' attention and interest to realize information level conversion, but they do not consider the influence of users' spatial location factors.

This study is jointly concerned with when and how to display content. These decisions are made based on the user's intentions. Different from the aforementioned adaptive method, what it considers is how to recognize the user's intention so that the display form of the virtual information can be automatically determined during the assembly process to visually enhance the information that the user needs.

2.3 User intention recognition in augmented reality

In the field of augmented reality, the recognition of user intention is mostly carried out in the form of natural human–computer interaction. Natural gesture operation is one of the most direct expression methods for users to convey their cognition. AR system can understand the user's behavior by capturing the user's natural gestures, so as to perceive the user's intention and make corresponding response behavior [45]. In addition, eye tracking, as an accurate attention indicator, can tell us what we are interested in [46]. Cognitive science and human–computer interaction (HCI) research have demonstrated that eye tracking can be used to measure user intention, characteristics, and status, and provide active and passive input control for AR interface. Sensing the gaze point and gaze time through the sensor can dynamically track the user's intention and state [30]. Gaze has the potential to provide the AR interface with the advantage of avoiding information overload and providing information on demand [29]. McNamara A et al. [46] realized the technology of placing information tags in a complex virtual environment (VE) by using eye tracking as an accurate attention indicator to capture objects of interest to the user. When the user noticed these objects of interest, the

system would display the tags associated with them to reduce information overload. Pfeuffer et al. [29] designed a gaze interaction design space to capture the user's visual attention through the gaze interaction dwell time in the AR environment. In the process of presenting multi-layer information, the system can transition from the initial information to gradually more detailed information through the user's gaze dwell time, so as to achieve a wider range of user information needs. In addition to gesture and gaze interaction, user intention can also be perceived through multi-channel interaction [47]. Seipel et al. [48] proposed a method of using natural language understanding technology to perceive the user's intention, which can enable users to make use of both gesture and speech actions simultaneously for information exploration. Their system also considered the contextual information of the visualization and made the conversational interface based on utterances "automatically" followed the conversation and adapted the visualization accordingly. Pfeuffer K et al. [49] explored a special combination of interactive technology in virtual reality, that was gaze-squeeze interaction technology, which used eye gaze to select targets and gestures to manipulate targets. This method perceived the user's intentions through hand-eye collaborative interaction, and realized a series of advanced interactive 3D operations on targets at any distance in the virtual environment.

In an assembly site with a complex and noisy environment, voice interaction will be affected by noisy sounds. In addition, the user needs two hands to operate the physical part. Therefore, it is not suitable to perceive the user's intention by voice and gesture. Perceiving the user's intention through eye gaze mentioned above can be used as a pre-processing step to this approach. Different from the aforementioned work, to our best knowledge, this paper comprehensively considers the user's eye gaze information, the user's spatial position, and the user's current assembly task in AR assembly for the first time. By quantifying these factors, this paper can perceive the user's intention and provide the virtual information that the user may need to reduce the user's cognitive burden.

3 Methodology

This section presents the design and detail of the User-Intention Adaptive Visualization System (UIAVS). UIAVS considers the user's intention to adaptively change the visual form of guidance information. The prototype is described with implementation specifics covering (1) system framework, (2) content creation, (3) user intention recognition, and (4) adaptive visualization.

3.1 System framework

The prototype system includes three main modules: (1) a content creation client that provides process designers with a platform to generate AR instructions; (2) an AR visualization client supports adaptively adjusting the visual form of AR assembly instructions based on user intention; (3) a server client connects the content creation client with the AR visualization client to manage and share task process resources.

Figure 1 presents the overall workflow of the User-Intention Adaptive Visualization System (UIAVS). The goal of this research is to automatically adjust the visibility, level of detail (LOD), and visual form of each virtual element in an AR assembly environment according to the user's intention. On the content creation client, the process designer can import the 3D CAD model of the assembly and generate AR assembly instructions. These instructions are composed of virtual elements, such as text, images, 3D CAD models, and assembly animations, which differ from traditional assembly processes. Process designers can set the initial parameters that are required to generate AR assembly instructions before using the system, and define the trigger conditions for variable parameters that may change during the system operation. Variable parameters include the visibility, the level of detail (LOD) display priority order, and the visual form triggers for each virtual element in the current task. The system can dynamically adjust variable parameters based on user intention during operation. Then, these AR assembly instructions are uploaded to the server client by the process designer. The server client performs data storage, analysis, processing, and transmission tasks between the content creation client and the AR visualization client. Moreover, it gathers user behavior information from the AR visualization client, such as user spatial location and eye gaze. Using the predefined logic rules, it deduces user intention from this information. Next, it synthesizes and analyzes this information with the assembly instruction information from the content creation client to generate AR assembly instruction logic data that can change the visual form of virtual elements. Lastly, it delivers the AR assembly logic data to the AR visualization client. The AR visualization client includes the AR visualization library and generates and visualizes AR assembly instructions. It receives the AR assembly instruction generation logic and the 3D CAD model of the part from the server client and converts them into AR assembly instructions that can change the visual form of virtual elements. The AR assembly instructions on the AR visualization client can dynamically change the visual form of virtual elements based on user behavior, to provide the necessary information and instruct the user to complete assembly tasks.

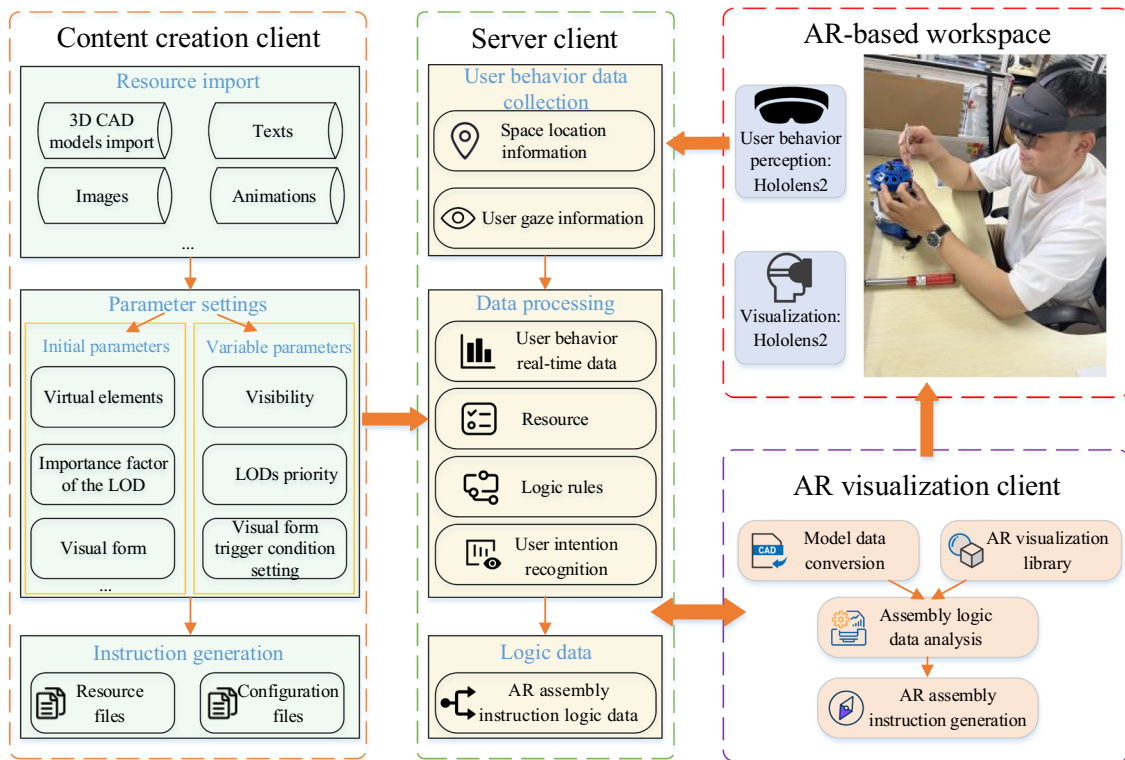


Fig. 1 The workflow of the User-Intention Adaptive Visualization System

The prototype system was developed using the Unity3D 2021.1.4f1c1 game engine and Microsoft's Mixed Reality Toolkit (MRTK) on Windows 10 operating system (see Fig. 2). MRTK is an open-source augmented reality development toolkit that provides a unified input system that supports various input devices and modes, such as gestures, eye tracking, and voice, and enables developers to rapidly create cross-platform MR applications. The content creation client used a Dell Alienware 17 (ALW17C-D2758) laptop as its hardware platform. The laptop was equipped with an Intel Core i7 7700HQ 2.8 GHz CPU, an NVIDIA GeForce GTX 1070 graphics card, and 16 GB of memory. WiFi network connection was also available on the laptop. The hardware platform for the AR visualization client was HoloLens 2, which has the ability to collect and display data. HoloLens 2 is an augmented reality headset from Microsoft with features such as spatial awareness, eye tracking, gesture recognition, and voice control. The server client used an Intel NUC717BNH microcomputer as its hardware platform. The microcomputer was equipped with an Intel Core i7 7567U 3.5 GHz CPU, an Intel GMA HD 650 graphics card, a Windows 10 professional 64-bit operating system, and 6 GB of memory. WiFi network connection was also available on the microcomputer.

On the content creation client, the user who plays the role of process designer can control the visibility, the LOD

importance division, and the visual form trigger conditions of each virtual element. The process designer imports the assembly 3D CAD models into the Unity 3D game engine and generated prefabs and asset bundles resource files. The process designer also creates guidance information LODs panels based on the assembly process and other related information and converted them into resource files in Unity 3D. Then, the process designer packages and submits the resource files to the server client through WiFi. The configuration files, including the file of the assembly logic data, can also be defined by the process designer. Unity 3D can parse the assembly logic data into an XML format file that organized the assembly logic into a tree hierarchy supporting AR instruction representation. The XML format file can then be packaged and submitted to the server client by the process designer. For details of this work, please refer to Ref [50]. Different from the previous work, this study has designed a new interface so that process designer can freely set the visibility of virtual elements, the display priority order of LODs, and the trigger conditions of various visual forms.

On the server client, data processing, communication, and data sharing are performed by the server among the content creation client and the AR visualization client. The server client receives the user's spatial position and eye gaze data collected by HoloLens2. It analyzes and processes these data, recognizes the user's intention based on predefined logic

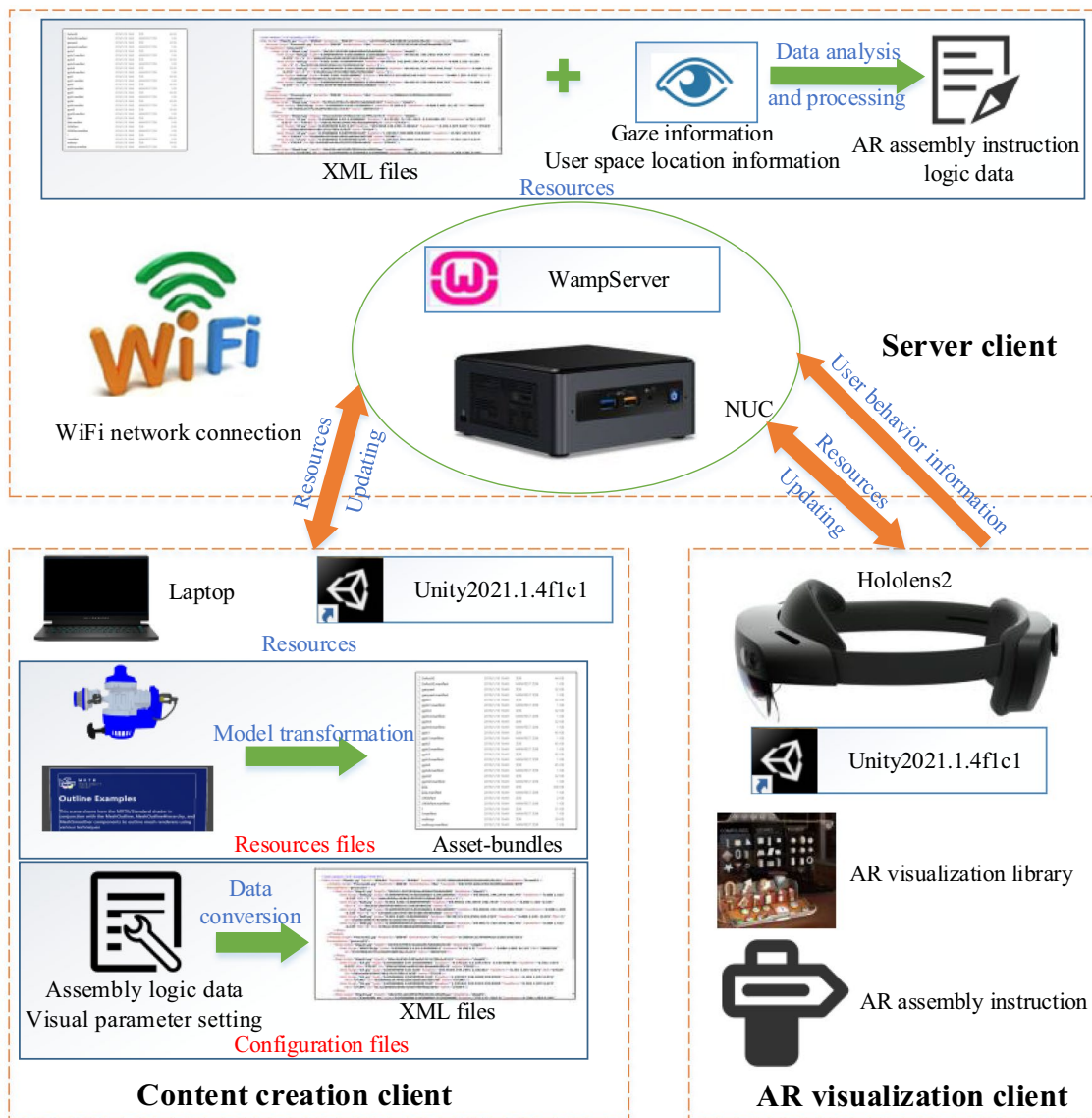


Fig. 2 The framework of the User-Intention Adaptive Visualization System

rules, and links them with the 3D CAD models, configuration files, LODs panels, and other information from the content creation client (see Sect. 3.3 for more details). The server client processes information to form AR assembly instruction logic data that can change the visual form of virtual elements. It sends these data and the 3D CAD model of parts to the AR visualization client through the network using WampServer.

On the AR visualization client, the user who plays the role of the worker performs assembly tasks wearing Hololens2 head-mounted display. The AR visualization client has an AR visualization library for generating and visualizing AR assembly instructions. To unify the coordinate system between the virtual space and real-world space, this study used the method described by Piumsomboon et al. [51] for

virtual-real registration and calibration. The worker’s spatial position and gaze information are collected by Hololens2 and transmitted to the server client of the system. Then, the AR visualization client parses the AR assembly instruction logic data and the 3D CAD model of the part received from the server client into AR assembly instructions. AR assembly instructions can recognize the user’s intention based on the user’s spatial position and gaze information and adaptively change the visualization form of virtual elements. Finally, the worker can follow AR assembly instructions to perform assembly tasks. It should be noted that the 3D CAD model of the part only needs to be loaded at the initial startup, while the generation logic of the AR assembly instruction is loaded in real time during the assembly process.

Table 1 Initial parameters input supplied by process designers

Parameter	Description
$E = (e_1, \dots, e_n), n \in Z^+$	All n virtual elements in the current assembly task t
$D_e = (d_1, \dots, d_{m_e}), m_e \in Z^+$	Available m_e LODs for element e
$I_{e,d_e}^t \in [0, 1]$	Importance factor of the LOD $d_e \in D_e$ of element e
p_e^t	Usage frequency of element e during task t . The process designer can set a certain threshold P_e for p_e^t
u_{e,d_e}^t	Utility of e at the LOD d_e during task t . The process designer can set a certain threshold U_e for u_{e,d_e}^t
$VF_e = (1, \dots, vf_{m_e}), m_e \in Z^+$	Available visual forms of assembly part model e

Table 2 Variable parameters input determined by system

Parameter	Description
$V_e^t \in \{0, 1\}$	Determines if a virtual element e is visible during the current task t
$d_e \in D_e$	Determines virtual element e is displayed at which LOD
$vf_e \in VF_e$	Determines what visual form of the virtual element e is displayed

In summary, the system has the three key features: (1) it enables process designers to adjust the visibility, the LOD display priority order, and the visual form change trigger conditions for each virtual element according to their needs; (2) it recognizes user intentions by applying logical rules to user spatial position and eye gaze data; (3) it generates AR assembly instructions that adapt the visualization form of virtual elements (visibility, LOD, and visual form) to user intentions in real time. By considering user intentions, the system optimizes the visual form changes and provides the most relevant information for the current task. The next sections present the inputs and parameters required for content creation in the system, a user intention recognition method based on logical rules, and an adaptive visualization form adjustment approach.

3.2 Content creation

The system requires two types of inputs: initial parameters that are determined by the process designer before using the system (see Table 1) and variable parameters that are dynamically adjusted by the system at runtime according to the user’s intention (see Table 2).

3.2.1 Initial parameters input

For all virtual elements $e \in E = (e_1, \dots, e_n)$ in the current assembly task t , each element e requires a set of specifications before its visibility, LOD, and visualization form can be determined. This study uses a specification similar to the Ref [25]. $D_e = (d_1, \dots, d_{m_e})$ denotes a list that provides information about each virtual element e at different levels of detail, where m_e represents the number of specific level

information of element e (see Fig. 3). Each level of information, $d_e \in D_e$ has an associated importance factor I_{e,d_e}^t , which represents the importance of the LOD of element e to the user in the current task t . I_{e,d_e}^t determines the priority of d_e display. The importance factor I_{e,d_e}^t is determined by the process designer according to data-driven approaches [52]. Data sources include assembly process specification, workers’ experience information, and questionnaire information. For an individual task t , p_e^t represents the usage frequency of e . p_e^t is related to the visibility V_e^t of element e in assembly task t . The process designer can set a certain threshold P_e for p_e^t . It should be noted that when p_e^t reaches the threshold P_e , the display visibility priority of element e in the entire element set E can be changed. u_{e,d_e}^t denotes the utility of e at the LOD d_e . It represents the degree of use of d_e in the assembly process. The process designer also can set a certain threshold U_e for u_{e,d_e}^t . When u_{e,d_e}^t is greater than the threshold U_e , the display priority order of d_e in all LOD sets D_e can be changed. p_e^t and u_{e,d_e}^t can change by perceiving the user’s intention, which will be described in detail later. Note that although this method designs the display priority of LOD d_e in the set D_e , the process designer can still design more parameters to adjust the display order of d_e .

$VF_e = (1, \dots, vf_{m_e})$ denotes the list of various visual forms of assembly part model e , where m_e represents the number of variable visual forms for virtual model e . The trigger conditions for the transformation of visual form from one form to another are determined by the process designer. The system provides an easy AR visualization library for process designers. Process designers can directly change the visualization form by setting the number corresponding to the visualization form of AR visualization library, and set

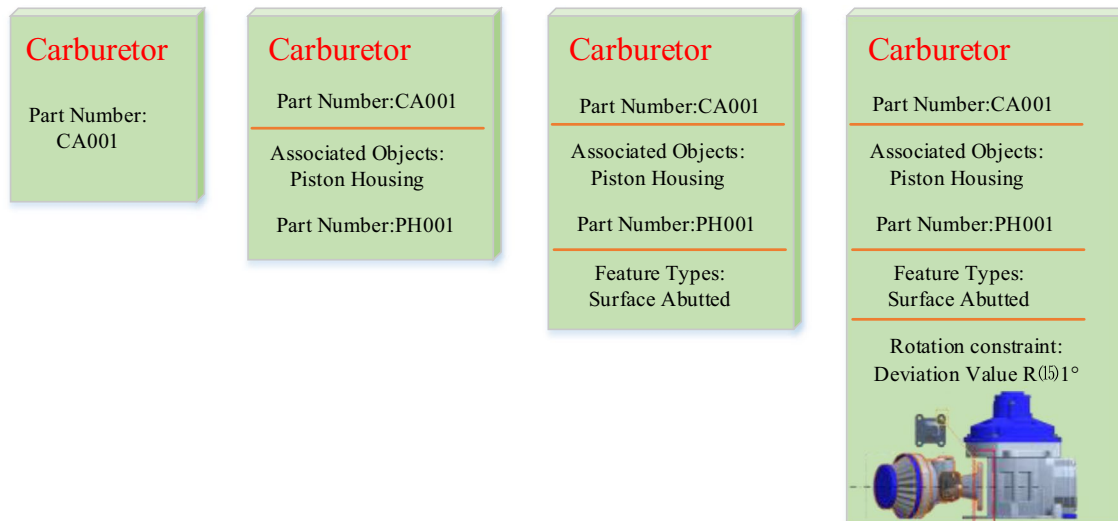


Fig. 3 Assembly process information of carburetor and its four LODs

simple parameters (such as resizing virtual elements and setting the transparency of virtual elements). The system also provides process designers with various threshold condition adjustment interfaces for triggering conditions of visual form changes of virtual models. The system then takes these as constraints into account. The system can adjust the visualization and LOD level of virtual elements without the user providing any explicit input. This manual procedure could be replaced with an automatic analysis.

It should be noted that the focus of this approach is not to automatically infer the user interaction required for the current task, nor to determine a mapping between the task and virtual elements e . There exist approaches for human activity recognition [53], which would plug into the system. When users feel that the visual form automatically adjusted by the system does not meet their needs, the system also provides users with an operation space for free interaction. The main objective of this research is to develop a flexible and general method of AR adaptive adjustment visualization to provide users with important and needed information. Connecting this with more advanced sources of information is a natural extension of this work.

3.2.2 Variable parameters input

Some initial parameters determined by the process designer are not invariable. They will be changed according to the user's intention when the system is running. The visibility V_e^t of a virtual element e during the current task t will be initially determined by the process designer. When the system is running, it will continue to determine whether the virtual element e is visible according to the user's spatial location and behavior. Similarly, the system will continue

to determine element e is displayed at which LOD. At the beginning, the process designer sets the priority display order of d_e according to its importance in the current task t , and sets the trigger conditions for the change of display order of d_e . When the conditions are met, the system will automatically decide element e to display at which LOD to provide important information for users. In addition, the visual form vf_e of the virtual element e is also a variable parameter. For most of the virtual elements that are indicative of information, the visual form does not change much, but for some virtual assembly models e , the change of the visual form helps users to understand the assembly process more easily. This study uses logical rules to recognize the user's intention and adjust the variable parameters accordingly.

3.3 User intention recognition

In the system, the adaptive process is to determine the user intention in the assembly process according to the user's behavior and automatically adjust the visualization form of virtual elements. The recognition of user intention is the key to adaptively adjusting the visualization form. Users' intentions are diverse and difficult to recognize. Therefore, it should be noted that the main focus of this study is the intention of users to adjust the visual form in order to better understand the assembly information during the assembly process.

3.3.1 User intention classification

In this method, the user's intention can be classified into three categories: (1) the visibility of the virtual element e can provide intuitive and important assembly information

in the current assembly task; (2) in the current assembly task, the required information can be obtained by searching the LOD of the virtual element e ; (3) the exploration of the virtual part model e can obtain an intuitive visual form, which makes it easier to understand the current assembly process. Based on this, the method is to perceive the user's behavior through hardware devices such as sensors, then process the perceived data, and set logic rules to analyze the user's intention.

3.3.2 User behavior perception

(1) Spatial location

Location-aware systems, also known as context-aware systems, have become increasingly important in recent years [28]. Users can express their interest through simple head posture and keep their hands free. The user's head position and head direction are obtained by the head motion sensor, and the area of interest to the user can be obtained. When the user approaches the destination, the visual form of the virtual element of the destination can be adjusted according to the change of the region of interest. The system uses a location tracking algorithm based on artificial markers to locate assembly parts, and scans the artificial marker through HoloLens2 to deeply integrate the virtual model with the real object in the assembly process. In the assembly process, the position of the virtual model of the part to be assembled is fixed and registered on the real object. Therefore, this study uses the distance parameter dt between the user's spatial location and the location of the currently assembled virtual part as one of the data sources for us to analyze the user's intention. The disadvantage of adaptively adjusting the visualization form only through the user's spatial position is that the change is single and does not provide the user with free interaction. Therefore, the system perceives the user's point of interest and the user's intention through eye tracking.

(2) Eye tracking

Eye tracking is a proxy for attention on specific AR information that can tell us what we are interested in [46]. Fixation point p is the basic output measure of interest, which can show what target the eye is looking at and select virtual elements. Gaze allows users to browse multiple information levels, from concise and important information levels to detailed levels, so as to provide users with a way to gradually consume more information. Users can complete the conversion of LOD information levels through gaze interaction. Deeper information transformation can unfold carefully according to the attention to specific node elements in the existing information level. To transition from initial

information to gradually more detail, the system should take the visual attention into account in the process of presenting multiple levels of information. This can be designed by temporal and spatial multiplexing of the virtual elements in relation to gaze data [29]. The most prominent selection method via gaze is the dwell-time mechanism, which is to gaze at an object over a specific amount of time. Timings for target selection can refer to the gazing design principle of HoloLens2.¹ Dwell-time threshold can be from 150 to 1500 ms [54, 55]. We are concerned about the gaze time t_g that can change the display level of LOD. First, this study directly provides the important information display level, and then gradually expands the display information level by gazing at the time t of virtual elements, while allowing us to adapt to the continuous development of information. The conversion of the visual form of the virtual part model is also similar to the LOD level change display mechanism. The difference is that what changes through the dwell-time mechanism is not the display level of LOD, but the change of visual form. In addition, the system counts the total time t_{sum}^e and the total number of times n_{sum}^e of gazing at element e in assembly task t . When the total time t_{sum}^e or the total number of times n_{sum}^e reaches a critical value set by the process designer, the LOD level and visualization form also change accordingly. Therefore, the fixation point p and gaze time t_g , as well as the total gaze time t_{sum}^e or total gaze number of times n_{sum}^e of a virtual element e , in the current task are also used as the data sources for us to analyze the user's intention.

3.3.3 Intention recognition based on logical rules

In assembly scenarios, user intention recognition can be achieved by logic rule-based methods that exploit rules derived or induced from domain knowledge [56]. Depending on the specific assembly task, a rule-based method can be used to infer the user's intention and provide adaptive visualization feedback based on the user's behavior and the assembly part information. As an example, suppose that the current assembly task involves screwing a bolt. The user grabs a bolt from the assembly part table and walks to the part that needs to be assembled. The rule-based intention recognition system infers that the user's intention is to find the correct position for the bolt on the part to be assembled and provides adaptive visualization feedback by highlighting that position. When the user approaches the assembly position, the system infers from the user's gaze information that the user wants to know the required torque for tightening the bolt and displays that information with high priority. This study conducts a comprehensive analysis of how

¹ <https://docs.microsoft.com/en-us/windows/mixed-reality/design/gaze-and-dwell-head>

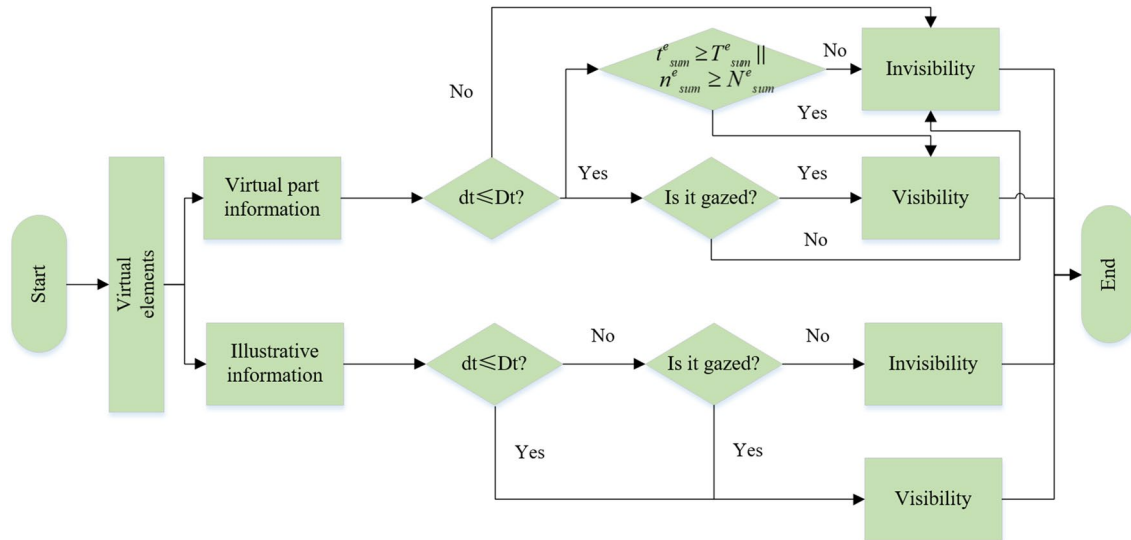


Fig. 4 The workflow of user intention recognition of adjusting the visibility of visualization information based on logical rules. dt represents the distance between the user’s spatial location and the position of the currently assembled virtual part. Dt denotes the threshold value

of the distance parameter. t^e_{sum} and n^e_{sum} represent the total duration and the total frequency of gazing at element e , respectively. T^e_{sum} and N^e_{sum} denote the threshold values of the total duration and the total frequency of gazing at the virtual element e , respectively

user behavior and assembly part visualization information interact. Each assembly task requires 3D CAD models of the parts to be assembled and the assembly process to be decomposed into virtual elements, which are all explicitly defined. Based on the user intention classification presented in Sect. 3.3.1. This study categorizes the user’s intention to change the visualization information of the virtual elements

during the assembly process into three types: (1) visibility, (2) LODs priority, and (3) visualization form. This study designs logical rules for user intention recognition from both spatial and temporal perspectives. This study defines logical rules that can accurately recognize the user’s intention to change the visualization information of the virtual elements based on the user’s spatial distance, gaze area

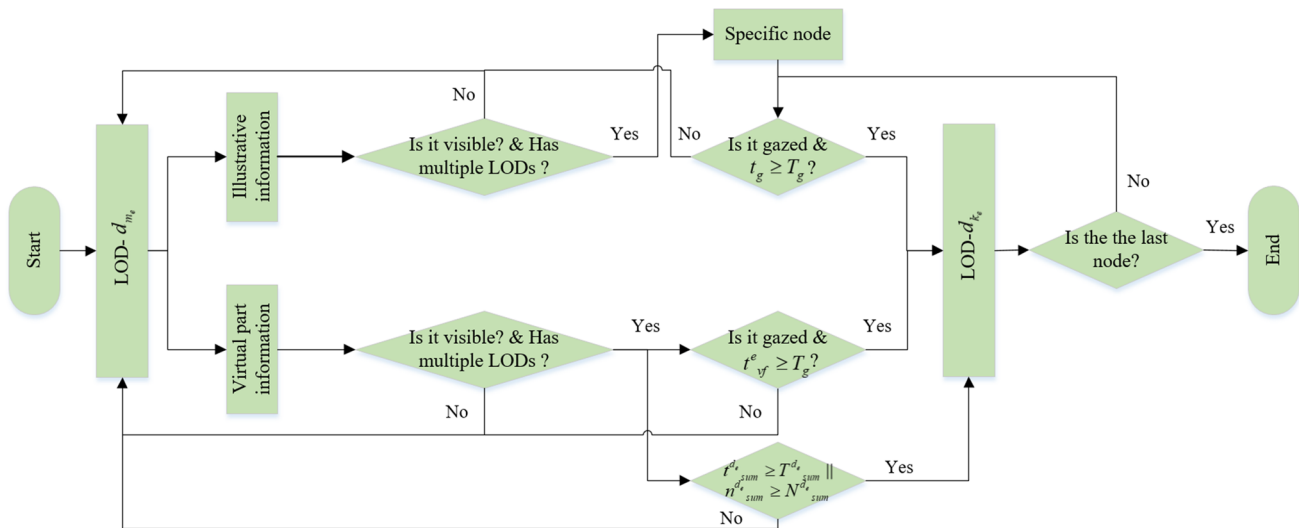


Fig. 5 The workflow of user intention recognition of adjusting the LODs priority of visualization information based on logical rules. d_{m_e} represents the current LOD level, and d_{k_e} represents the next level of information to be displayed. t_g represents the duration of gazing at the specific node of the LOD. t^{e}_{vf} represents the duration of gazing at

the virtual part e . T_g denotes the threshold value of the gaze dwell time. $t^{d_e}_{sum}$ and $n^{d_e}_{sum}$ represent the total duration and the total frequency of gazing at the LOD level d_e , respectively. $T^{d_e}_{sum}$ and $N^{d_e}_{sum}$ denote the threshold values of the total duration and the total frequency of gazing at the LOD level d_e , respectively

and duration, and the threshold parameters specified by the process designer. By constantly updating and analyzing the user’s behavior parameters $B = \langle p, t_g, t_{sum}^e, n_{sum}^e, dt \rangle$, the system recognizes the user’s intention and changes the visual form of the virtual element. In the following, the logical rules for user intention recognition are described in detail.

(1) Visibility

For the convenience of description, the visibility of virtual elements is divided into the visibility of virtual part information $V_{vf_e}^t \in \{0, 1\}$ and the visibility of illustrative information $V_{d_e}^t \in \{0, 1\}$. Figure 4 illustrates the workflow of recognizing the user’s intention of adjusting the visibility of visualization information based on logical rules. The virtual model of the part to be assembled in the current assembly task t has been specified by the process designer to be in the highlighted visible state. The visibility $V_{vf_e}^t \in \{0, 1\}$ of other virtual parts e in the assembly is specified as follows:

$$V_{vf_e}^t = \begin{cases} 1, & (p_e = 1, dt \leq Dt) \\ 1, & (t_{sum}^e \geq T_{sum}^e \parallel n_{sum}^e \geq N_{sum}^e, dt \leq Dt) \\ 0, & (otherwise) \end{cases} \quad (1)$$

$p_e \in \{0, 1\}$ denotes whether the virtual part e is gazed. Dt represents the distance threshold set by the process designer. T_{sum}^e and N_{sum}^e are thresholds for the total duration and the total frequency of gazing at element e set by the process designer. When the user’s distance dt from the current assembly virtual part registered on the assembly is greater than the set threshold Dt , the current assembly virtual part is highlighted, and other parts are not displayed whether they are gazed or not. The gaze behavior of the real parts can be monitored by registering the virtual parts and the real assembly parts together and tracking the gaze behavior of the virtual parts. When dt does not exceed the set threshold Dt , other parts are highlighted when they are gazed. In addition, when t_{sum}^e or n_{sum}^e reaches the set threshold T_{sum}^e and threshold N_{sum}^e , it indicates that the virtual part e is the information that the user may need to pay attention to, so it is highlighted whether it is gazed or not.

The visibility of the most basic and important information such as the name and part number of the virtual part e is the same as the visibility $V_{vf_e}^t$ of the virtual part e . And the visibility $V_{d_e}^t \in \{0, 1\}$ of illustrative information of virtual part e in the assembly is specified as follows:

$$V_{d_e}^t = \begin{cases} 1, & (p_e = 1, dt > Dt) \\ 1, & (dt \leq Dt) \\ 0, & (otherwise) \end{cases} \quad (2)$$

When the distance dt between the user and the virtual part e is greater than the set threshold Dt , the LOD level information of the highest priority display set by the process designer is displayed only when the user gazes at the virtual part e . When the distance dt does not exceed the set threshold Dt , the most priority important information is displayed whether or not the user gazes at the virtual part e .

(2) Level of detail

After determining the visibility of the virtual element e , the goal of this research is to automatically determine at which level of LOD the illustrative information d_e is displayed to provide the information the user needs. Figure 5 illustrates the workflow of recognizing the user’s intention of adjusting the LODs priority of visualization information based on logical rules. The method is to change the LOD

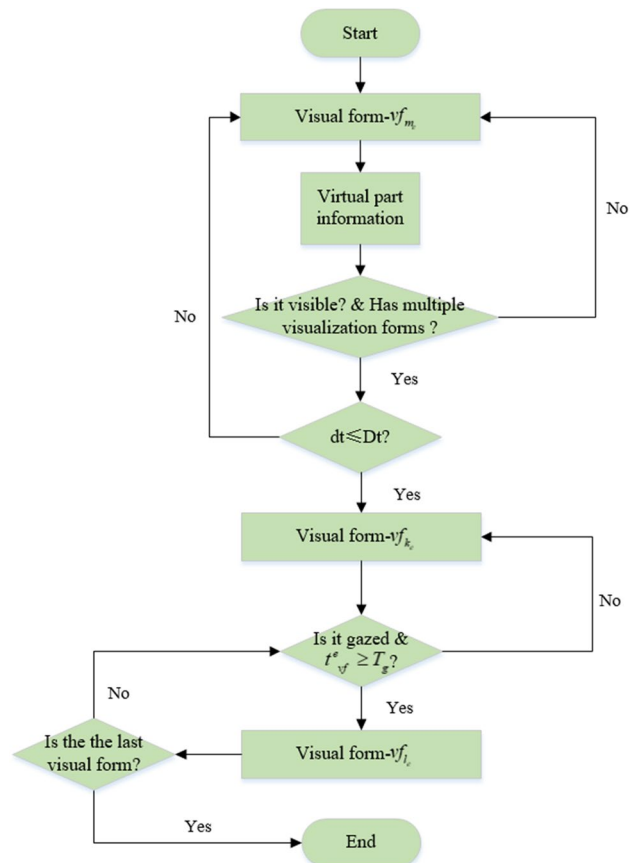


Fig. 6 The workflow of user intention recognition of adjusting the visual form of virtual parts based on logical rules. dt represents the distance between the user’s spatial location and the location of the currently assembled virtual part. Dt denotes the threshold value of the distance parameter. vf_m represents the current visual form of the virtual part e , and vf_k and vf_l represent the next and the subsequent visual forms to be displayed, respectively. t_{vf}^e represents the duration of gazing at the virtual part e . T_g denotes the threshold value of the gaze dwell time

through the dwell-time mechanism of gaze. If the user’s gaze time is too short, the automatic change of the LOD will cause the user to be disturbed. Therefore, a rule is set to fix the LOD. The LOD can only be changed when the user gazes at a specific node of the LOD and the gaze time t_g is greater than a threshold T_g (In this case is 0.8 s). The user gradually triggers the level by level change of LOD through gazing node and gazing duration. This can be formulated as:

$$\left. \begin{matrix} y_{d_e} = 1 \\ V^t_{d_e} = 1 \\ j_{d_e} = 1 \\ t_g \geq T_g \end{matrix} \right\} \Rightarrow d_{m_e} \rightarrow d_{k_e} \forall e \in E = (e_1, \dots, e_n), d_e \in D_e = (d_1, \dots, d_{m_e}, d_{k_e}, \dots, d_{n_e}) \tag{3}$$

$y_{d_e} \in \{0, 1\}$ denotes if virtual element e has multiple LODs. $j_{d_e} \in \{0, 1\}$ indicates whether the node is gazed or not. T_g denotes the gaze dwell-time threshold. d_{m_e} is the current LOD level, and d_{k_e} indicates the next level information to be displayed. $V^t_{d_e}$ indicates whether the illustrative information of the virtual part e is visible or not. It should be noted that in order to prevent excessive information display from causing trouble to users, the LOD display layer will gradually disappear after the user’s eyes look away from the LOD for a certain time. When the user looks at a certain threshold time again, the LOD display layer will be displayed again.

In addition, the user can also trigger the next level change of LOD when the time of gazing at the virtual part e is greater than a threshold T_g . This is mainly because the user may pay more attention to the virtual part e , so the system changes to display a more detailed LOD level to provide the user with more assembly information. The goal of this research is to display more common and important LOD level information. Therefore, when the total fixation time $t_{sum}^{d_e}$ or total fixation number $n_{sum}^{d_e}$ of a LOD level d_e reaches the set time threshold $T_{sum}^{d_e}$ or number threshold $N_{sum}^{d_e}$, the LOD display priority can be automatically changed. This can be formulated as:

$$\left. \begin{matrix} y_{d_e} = 1 \\ V^t_{d_e} = 1 \\ t^e_{vf} \geq T_g \end{matrix} \right\} \Rightarrow d_{m_e} \rightarrow d_{k_e} \forall e \in E = (e_1, \dots, e_n), d_e \in D_e = (d_1, \dots, d_{m_e}, d_{k_e}, \dots, d_{n_e}) \tag{4}$$

$$\left. \begin{matrix} y_{d_e} = 1 \\ V^t_{d_e} = 1 \\ t^e_{sum} \geq T^d_{sum} \parallel n^d_{sum} \geq N^d_{sum} \end{matrix} \right\} \Rightarrow d_{m_e} \rightarrow d_{k_e} \forall e \in E = (e_1, \dots, e_n), d_e \in D_e = (d_1, \dots, d_{m_e}, d_{k_e}, \dots, d_{n_e}) \tag{5}$$

t^e_{vf} indicates the duration of gazing at the virtual part e . $T_{sum}^{d_e}$ and $N_{sum}^{d_e}$ are the thresholds for the total duration and the total frequency of gazing at the LOD level set by the process designer.

(3) Visual form

The adjustment of the visual form is for the virtual part e . Figure 6 illustrates the workflow of recognizing the user’s intention of adjusting the visual form of virtual parts based on logical rules.

The goal of this research is to provide different visual forms of virtual parts at different stages of user assembly to attract users’ attention and make it easier for users to understand the assembly information of the current assembly step. In the AR visualization library, the visualization form can take various forms, such as magnification, transparency, or wireframe model. The priority order of visualization form adjustment is determined by the process designer by changing the number corresponding to the visualization form in the AR visualization library.

The method changes the visual form of virtual parts through the change of user’s spatial location and gaze dwell-time mechanism. For example, when the distance dt between the user’s spatial location and the position of the current assembly virtual part is greater than the set threshold Dt , the key information that the user pays attention to is not the current assembly virtual part itself, so its visualization form is not adjusted. When dt does not exceed the set distance threshold Dt , the current assembly virtual part is magnified (or other set visual form of priority display), and the magnification depends on the multiple set by the process designer. This can be formulated as:

$$\left. \begin{matrix} y^e_{vf} = 1 \\ V^t_{vf_e} = 1 \\ dt \leq Dt \end{matrix} \right\} \Rightarrow vf_{m_e} \rightarrow vf_{k_e} \forall e \in E = (e_1, \dots, e_n), vf_e \in VF_e = (1, \dots, vf_{m_e}, vf_{k_e}, \dots, vf_{n_e}) \tag{6}$$

$y_{vf}^e \in \{0, 1\}$ indicates whether the virtual part e has multiple visualization forms. vf_{m_e} is the current visual form of the virtual part e , and vf_{k_e} represents the next visual form to be displayed. $V_{vf_e}^t$ indicates whether the virtual part e is visible or not.

In addition, users can also adjust the visualization form through the gaze dwell-time mechanism. When the time t_{vf}^e

$$\left. \begin{array}{l} y_{vf}^e = 1 \\ V_{vf_e}^t = 1 \\ dt \leq Dt \\ t_{vf}^e \geq T_g \end{array} \right\} \Rightarrow vf_{k_e} \rightarrow vf_{l_e} \forall e \in E = (e_1, \dots, e_n), vf_e \in VF_e = (1, \dots, vf_{k_e}, vf_{l_e}, \dots, vf_{n_e}) \quad (7)$$

vf_{k_e} is the current visual form of the virtual part e , and vf_{l_e} represents the next visual form to be displayed. It should be noted that the change of the visual form of the virtual part e is not independent, and the level of its LOD will also change as described in the previous section. In addition, this study also provides an explicit interaction method, which allows users to freely interact and choose other visualization forms when the current visualization form does not meet the requirements.

3.4 Adaptive visualization

According to the adaptive visualization rules formulated above, the user's intention is determined by comprehensively analyzing and processing the user's behavior parameter $B = \langle p, t_g, t_{sum}^e, n_{sum}^e, dt \rangle$. This study collects the user's behavior parameters using the head-mounted AR device HoloLens2 and uploads them to the server client in real time. Based on logical rules, this study identifies the user's intention from their behavior and derive variable parameters for visibility, level of detail (LOD), and visualization form (see formulas (1)–(7)). The server links these parameters with the resources created by the content creation client and produces logic data that can change the visualization form of virtual elements. These data and resources are then transmitted to the AR visualization client. The AR visualization client interprets them and integrates them with its local AR visualization library to generate AR assembly instructions. AR assembly instructions contain some parameter configurations that can change the visualization form of virtual elements. The system adjusts the visibility of virtual elements, the priority display level of LOD, and the visualization form of virtual parts in real time according to AR assembly instructions. It should be noted that AR assembly instructions change in real time according to the user behavior parameters. Therefore, the system can adaptively change the visualization form of virtual elements according to the user intention to provide the information needed by the user. Figures 7, 8, and 9 illustrate the effect of adaptive visualization of virtual element information.

of the user gazing at the virtual part e is greater than the set time threshold T_g , the virtual part e becomes a transparent display (or other priority change visualization methods set by the process designer). The user gradually triggers the layer by layer change of the visual form of the virtual part through the gazing time. This can be formulated as:

Figure 7 shows the effect of adaptive adjustment of visibility and information display of virtual elements based on user intention. When the user is in the part placement area, only the current assembly virtual part is visible (see Fig. 7a). However, in the part assembly area, other parts become visible when gazed at in the nearby region (see Fig. 7b). In addition, parts that are gazed at for a long time or frequently remain visible even when not gazed at in the nearby region (see Fig. 7c). It is worth noting that when the user is in the part placement area, only basic information of the current assembly virtual part is shown (see Fig. 7d). However, when the user gazes at it, the most important LOD information is displayed (see Fig. 7e). Furthermore, when the user is in the near-distance part assembly area, the most priority information is displayed regardless of whether they gaze at it or not (see Fig. 7f).

Figure 8 shows the effect of adaptive adjustment of LOD display priority of virtual elements information based on user intention. The user triggers the hidden nodes by eye gaze in the part assembly area and the LOD information can be displayed level by level (see Fig. 8a–c). Moreover, illustrative information can be displayed by the user's gaze at the virtual part in the same area (see Fig. 8d). Furthermore, after the user gazes at the virtual part for a certain time, another level of LOD information can be displayed (see Fig. 8e). Finally, the display priority of LOD can be updated automatically based on the user's gaze duration and frequency (see Fig. 8f).

Figure 9 shows the effect of adaptive adjustment of visualization forms of virtual elements information based on user intention. When the user is within the part placement area, the assembly position of the current virtual part can be effectively highlighted and visually represented, as depicted in Fig. 9a. Furthermore, as the user moves into the near-distance part assembly area, both the current assembly virtual part and its associated components can undergo a transformation in their visualization, such as an enlargement, as illustrated in Fig. 9b. Moreover, when the user directs gazes at the virtual part, its visual representation can

Fig. 7 The effect of adaptive adjustment of visibility and information display of virtual elements based on user intention

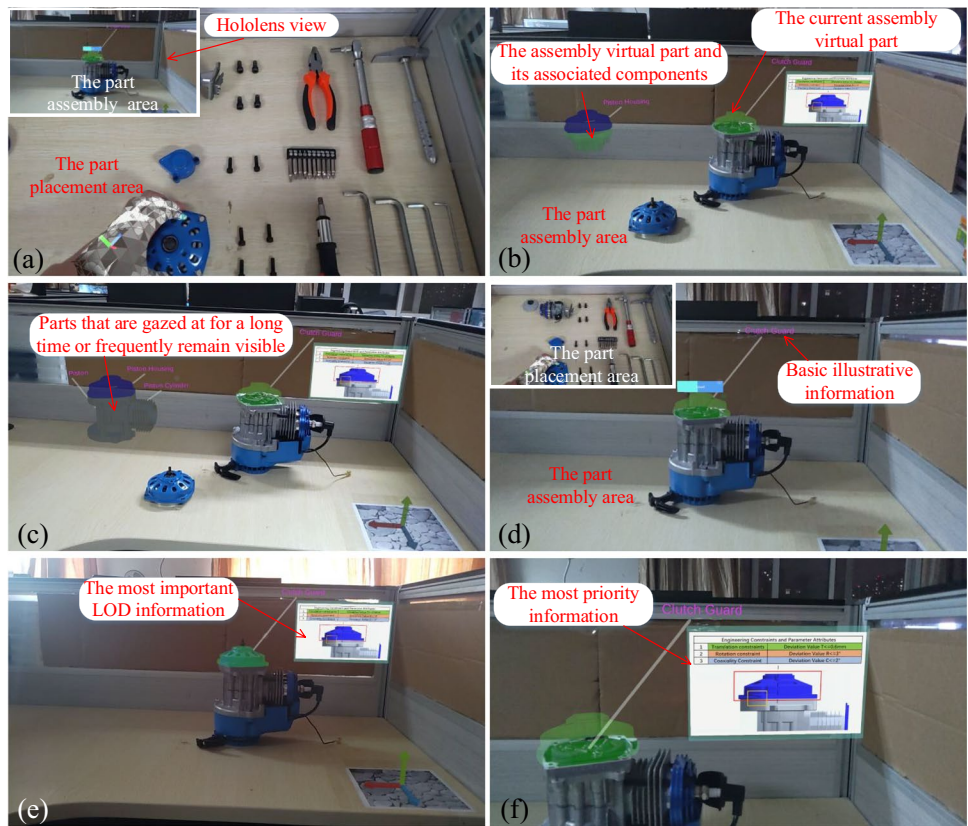


Fig. 8 The effect of adaptive adjustment of LOD display priority of virtual elements information based on user intention

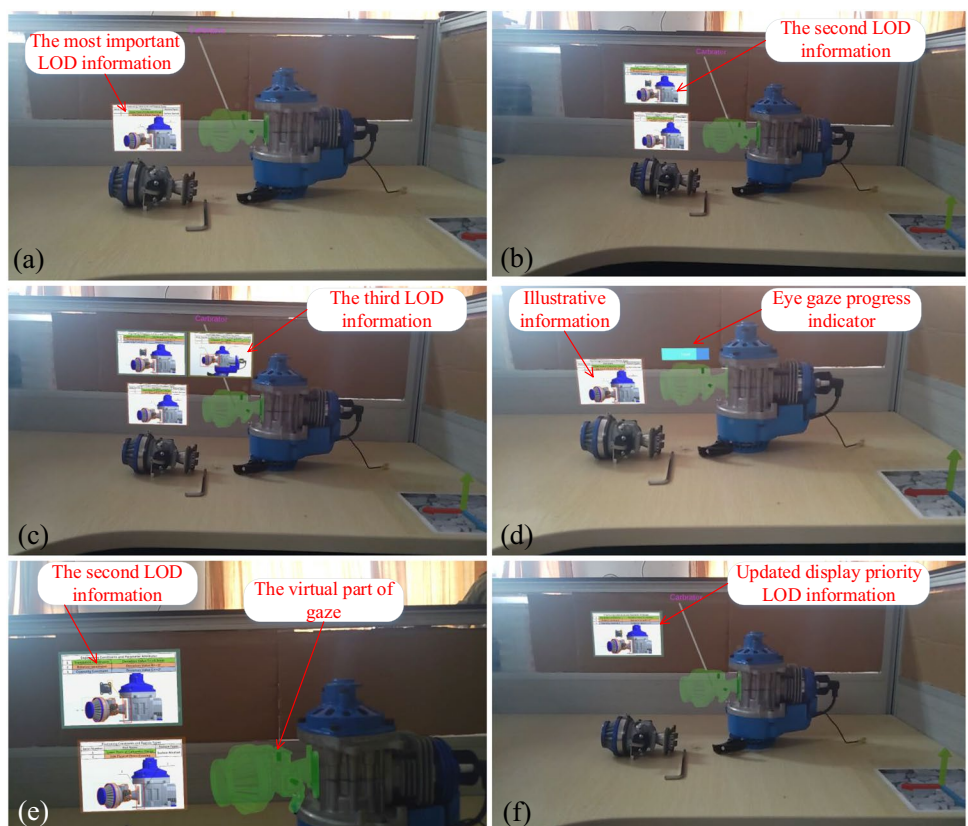
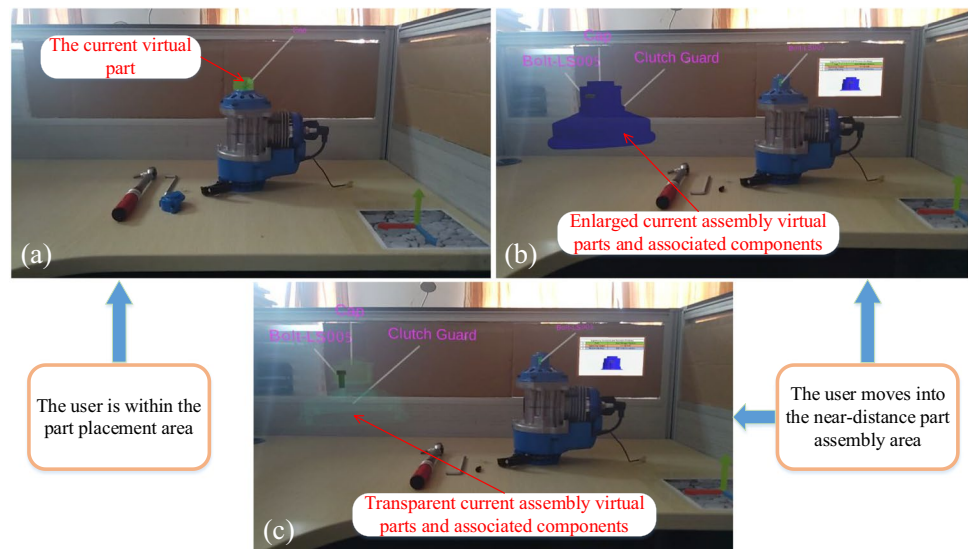


Fig. 9 The effect of adaptive adjustment of visualization forms of virtual elements information based on user intention



undergo sequential changes, such as rendering the related parts in a transparent manner, as shown in Fig. 9c.

4 User study

This section reports a user study on UIAVS to investigate the benefits and limitations of adaptive visualization method based on user intention, and to evaluate the effectiveness of the method in AR assembly process. In terms of research questions, we were interested in (1) how this method affects task performance and (2) how the system affects user attention and distraction. Considering the actual assembly situation, there is a difference in the cognitive level of assembly tasks between novice and expert. They may have differences in the information they need during the assembly process, so a grouped experimental design was performed. Therefore, the purpose of this study was to examine the effects of information visualization response methods on users' cognitive load, user experience, and task effectiveness, in two groups with different levels of task familiarity in assembly.

4.1 Study design

In this research, two information visualization response methods were selected:

- (1) **Baseline:** The most common visual response method in AR assembly. All information is directly displayed on the AR interface. The user manually adjusts the information visualization form.
- (2) **UIAVS:** An information adaptive visualization response method based on user's intention for AR assembly.

This user study used the between-subject design. The independent variables were information visualization response methods (Baseline or UIAVS) and level of expertise (novice or expert). The combination yielded 4 conditions, *Novice Baseline (NB)*, *Expert Baseline (EB)*, *Novice UIAVS (NU)*, *Expert UIAVS (EU)*. Dependent variables included task completion time, number of assembly errors, cognitive load, and user experience. Subjective cognitive load was measured by the NASA-TLX questionnaire [57], while user experience was evaluated by a self-designed user experience questionnaire with reference to [58, 59]. NASA-TLX and the user experience questionnaire were collected after experimenters completed their assembly task.

4.2 Setup and task

The experimental space was set up in a room (6.1 m by 4.3 m) that contained a simplified assembly table and a part placement table (see Fig. 10). The task was to locate the necessary components from the parts placement table and complete the assembly of a miniature engine on the assembly table.

The main parts of the engine were placed on the assembly table. Some target engine parts and tools were placed on the parts placement table. The users needed to find the components for assembly and the corresponding tools to complete the engine assembly task. This study chose this assembly task to simulate the process that users need to find parts and the final assembly position of parts to complete the assembly in the actual assembly. In such situations, users have to consciously make an effort to find information in the interface during AR assembly. Depending on the visible virtual elements and their LOD, the users need to perform interactions

Fig. 10 The setting of the experimental work scenario. **a** The part placement area. **b** The part assembly area

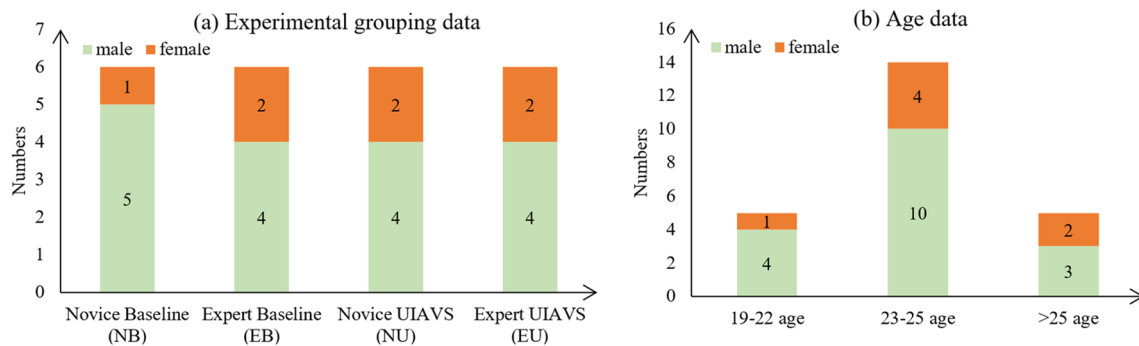
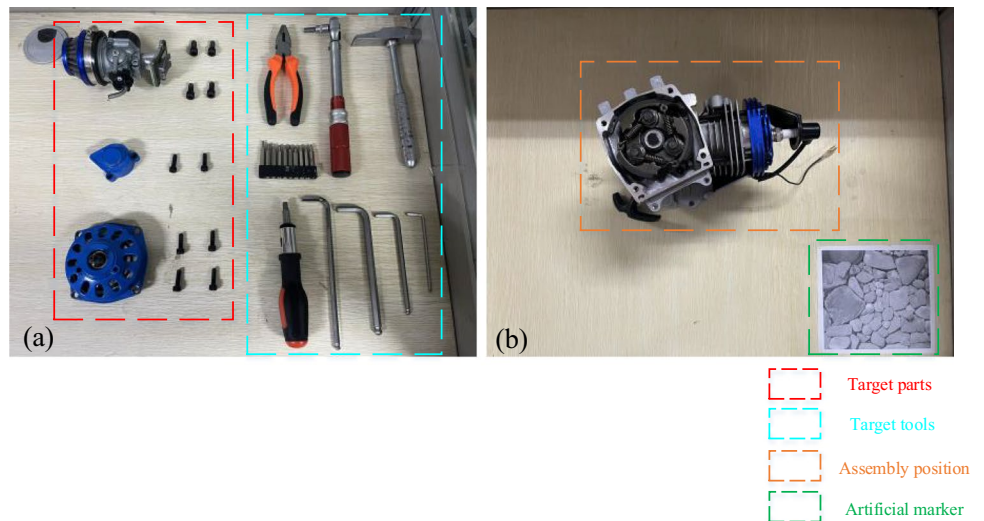


Fig. 11 Statistical data of participants in the experiment. **a** Distribution data of participants in four experimental groups. **b** Age data of experiment participant participants

to find information and complete the task according to the instructions.

4.3 Hypotheses

Chanquoy et al. [60] explained that the higher the professional level and familiarity in the task, the lower the load. In the actual assembly, for the same assembly task, the expert is more familiar than the novice. Therefore, the expert may have a lower cognitive load than the novice, and the amount of information required in the AR assembly may be less. In addition, the studies [61, 62] showed that both system properties (such as information visualization response method) and user characteristics (such as task familiarity level) have effects on human–computer interaction. These interactions impact the perception of practicality and focus, and then affect the user experience. Based on this and earlier research results, the following hypotheses were proposed:

H1: Time. The UIAVS interface will be more efficient than the Baseline interface in task completion time.

H2: Error. Using UIAVS will reduce operating errors.

H3: User cognitive load. The cognitive load of users using UIAVS is lower than that of users using Baseline, whether novice or expert.

H4: User Experience(UX). UIAVS will provide a better user experience than the Baseline.

H5: Number of interactions. The number of user interactions in UIAVS is lower than in Baseline.

4.4 Participants

Twenty-four participants (17 males and 7 females, mean age of 24.3 years, SD = 2.9) from Northwestern Polytechnical University were invited for this study (see Fig. 11). We sought participants with AR/VR experience to reduce the impact of novelty effects. They were randomly and evenly assigned to one of two information visualization response

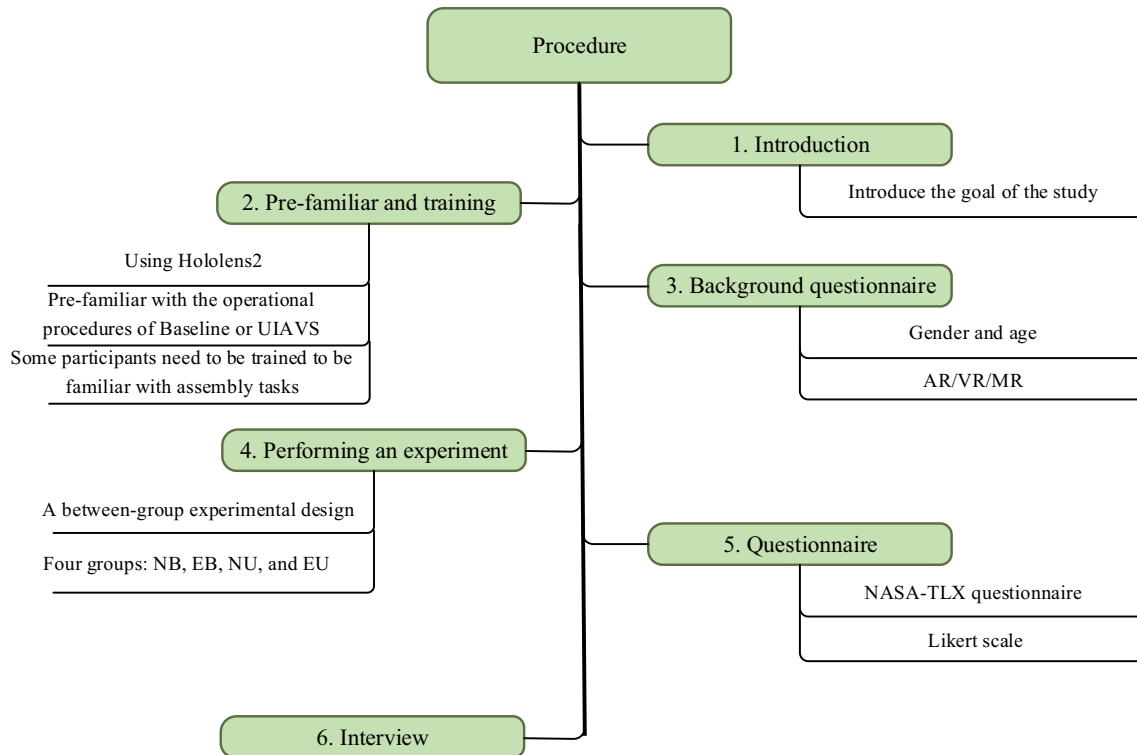


Fig. 12 The procedure of user study

methods (UIAVS or Baseline) and one of two levels of task familiarity (novice or expert).

Four experimental groups were constituted. Twelve participants assigned to the expert level group were trained to become familiar with the assembly task. They are familiar with the assembly process but cannot remember all the key data. This was to imitate the actual situation of most workers in the assembly process.

4.5 Procedure

The user study followed the six steps as shown in Fig. 12. Participants were informed of the goal of the experiment. Participants needed to search for information in the AR interface to find the information required for the current assembly task as soon as possible. In addition, participants would be required to be pre-familiar with the operational procedures of Baseline and UIAVS. Researchers would then explain the meaning of each data parameter to participants and spend as much time as possible allowing participants to fully understand the content of the instructions provided. Before the formal experiment, each participant was required to complete a brief research background questionnaire. In the experiments, four groups were divided for a between-group experimental design, namely *NB*, *EB*, *NU*, and *EU* (see Fig. 11). During the

experiment, participants were required to complete the assembly of the engine. Figure 13 shows the main steps of the engine assembly task using the UIAVS interface by the participants. A timer was used to record the time each participant took to complete each task. The number of assembly errors for each participant was recorded after each task was completed. The system automatically recorded the number of interactions that each participant completed each task. At the end of task, participants were asked to fill out a NASA-TLX questionnaire and subjectively rated the user experience quality of each experiment on a scale of 1 (I totally disagree) to 7 (I totally agree). Each participant was asked to conduct interviews according to the content of the experiment.

4.6 Results

This section summarizes the data analysis results collected from five experimental measures: (1) performance time, (2) error evaluation, (3) cognitive load, (4) user experience, (5) number of interactions. Normality assessments on the data of all dimensions were conducted, and the homogeneity of variance was assessed through Levene's test. The between-groups analysis of variance was used when the data met the normality assumption and

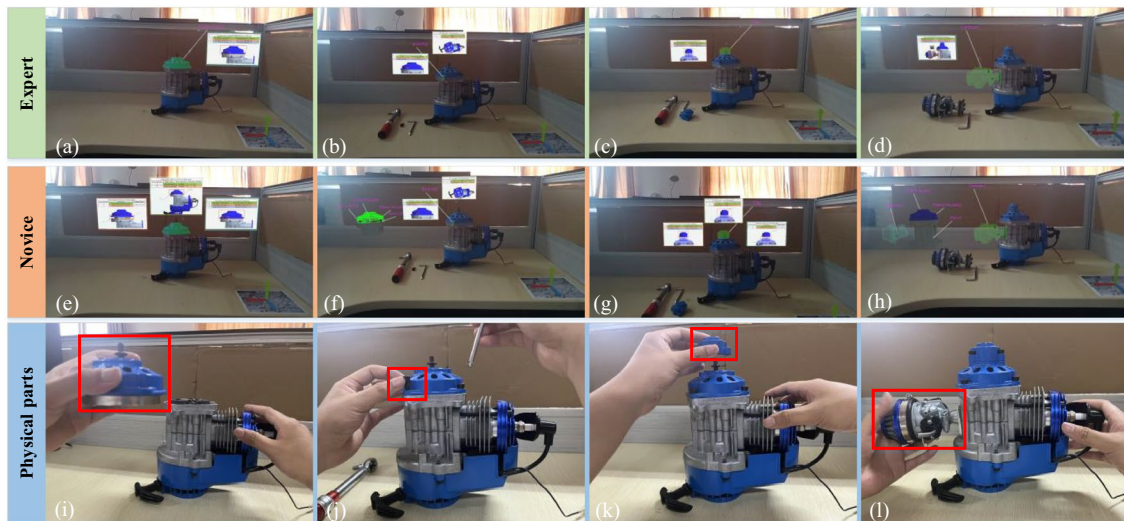


Fig. 13 The main steps of engine assembly using the UIAVS interface by the participants. **a–d** The Hololens view of the expert. **e–h** The Hololens view of the novice. **i–l** The physical parts assembly

Table 3 Assembly performance data of users according to visual response method and task familiarity level

	Novice		Expert		<i>F</i>
	Baseline	UIAVS	Baseline	UIAVS	
Performance time(s)	419.167 ± 18.986	372.667 ± 17.997	350.667 ± 17.143	235.333 ± 26.544	$F_v = 77.817^{**}$ $F_e = 125.883^{**}$ $F_{v \times e} = 14.078^*$
Number of errors	1.167 ± 0.687	1 ± 0.817	0.833 ± 0.687	0.667 ± 0.471	$F_v = 0.303$ $F_e = 1.212$ $F_{v \times e} = 0.001$
Cognitive load	15.154 ± 1.015	13.072 ± 0.977	10.542 ± 1.281	6.465 ± 0.630	$F_v = 47.161^{**}$ $F_e = 156.484^{**}$ $F_{v \times e} = 4.953^*$
Number of interactions	37 ± 1.291	29 ± 2.769	23.667 ± 1.886	9.167 ± 1.462	$F_v = 168.438^{**}$ $F_e = 365.998^{**}$ $F_{v \times e} = 14.057^*$

* indicates $p < 0.05$; ** indicates $p < 0.001$. F_v indicates the F -value of main effect of information visualization response methods. F_e indicates the F -value of main effect of level of expertise. $F_{v \times e}$ indicates the F -value of the interaction effect of information visualization response methods and level of expertise

homoscedasticity, and the nonparametric Mann–Whitney U test was used when the normality assumption or homoscedasticity was not met.

4.6.1 Performance time

The UIAVS interface and the Baseline interface were compared in terms of performance time in assembly tasks, with the aim of exploring the efficiency of the UIAVS interface. Table 3 shows the average performance time under different conditions. A two-way between-groups analysis of variance revealed the main effect of information visualization response methods on performance time was significant

($F(1,23) = 77.817, p < 0.001$). According to single-variable test, the average time to complete the assembly task using UIAVS ($M = 304.000$) interface was significantly shorter than that of Baseline ($M = 384.917, p < 0.001$), which was 21.0% shorter. There was a statistically main effect for level of expertise ($F(1,23) = 125.883, p < 0.001$). The single-variable test revealed that the average time to complete the assembly task for the expert ($M = 293.000$) was significantly shorter than that of the novice ($M = 395.917, p < 0.001$). The interaction effect between information visualization response methods and level of expertise was statistically significant ($F(1,23) = 14.078, p < 0.05$). Simple effect analysis showed that the average time ($M = 372.667$)

Table 4 Likert scale rating questions for user experience

Question	Question statement	Evaluation factors
Q1	This visual interface allows me to complete tasks quickly	Efficiency
Q2	I enjoyed this visual interface	Feeling
Q3	I can focus on my task without being distracted	Attention
Q4	I can easily find the information I need	Perception
Q5	The visual information provided is intuitive and useful	Usability
Q6	The AR instruction of this visual interface was helpful	Feasibility
Q7	I am confident that I have finished the task correctly	Confidence
Q8	This visual interface can understand my intention	Intention

for novices to complete assembly tasks using UIAVS was 11.1% shorter than that of Baseline ($M = 419.167$, $p < 0.05$). For the expert, the average time ($M = 235.333$) of using UIAVS to complete the assembly task was shorter than that of Baseline ($M = 350.667$, $p < 0.001$), which was 32.9% shorter.

4.6.2 Error evaluation

The initial hope of this research was to explore whether the use of UIAVS in assembly tasks could reduce the assembly error rate. However, to our surprise, according to the two-way between-groups analysis of variance, the interaction effect between information visualization response methods and level of expertise was not statistically significant ($p > 0.05$). There was no statistically significant main effect for either information visualization response methods ($p > 0.05$) or level of expertise ($p > 0.05$).

4.6.3 Cognitive load

Cognitive load is an important indicator to measure the effectiveness of information visualization response methods. NASA-TLX was used to measure cognitive load. The impact of information visualization response methods and level of expertise on global cognitive load was explored through a two-way between-groups analysis of variance. The results showed that there was a statistically significant main effect for information visualization response methods ($F(1,23) = 47.161$, $p < 0.001$). For users, compared with UIAVS interface ($M = 9.769$), Baseline interface would bring heavier cognitive load ($M = 12.848$, $p < 0.001$). There was a statistically main effect for level of expertise ($F(1,23) = 156.484$, $p < 0.001$). The single-variable test revealed that the cognitive load of the expert ($M = 8.504$) was significantly shorter than that of the novice ($M = 14.113$, $p < 0.001$). The interaction of information visualization response methods and level of expertise had a significant effect on the global cognitive load ($F(1,23) = 4.953$, $p < 0.05$). Simple effects analysis showed that novices using the

UIAVS interface ($M = 13.072$, $p < 0.05$) had a lighter cognitive load compared to Baseline interface ($M = 15.154$). In addition, for expert, the cognitive load of using the UIAVS interface ($M = 6.465$) was significantly lower than that of Baseline interface ($M = 10.542$, $p < 0.001$).

4.6.4 User experience

User experience is critical to the usability of a system. A 7-point Likert scale (see Table 4) was used to evaluate the user experience effect of the two information visualization response methods. Participants were asked to fill out a questionnaire that included 8 questions to assess efficiency (Q1), feeling (Q2), attention (Q3), perception (Q4), usability (Q5), feasibility (Q6), confidence (Q7), and intention (Q8). The Mann–Whitney U test was used to explore whether there was a difference in user experience between the UIAVS interface and the Baseline interface. The statistical data results are shown in Fig. 14.

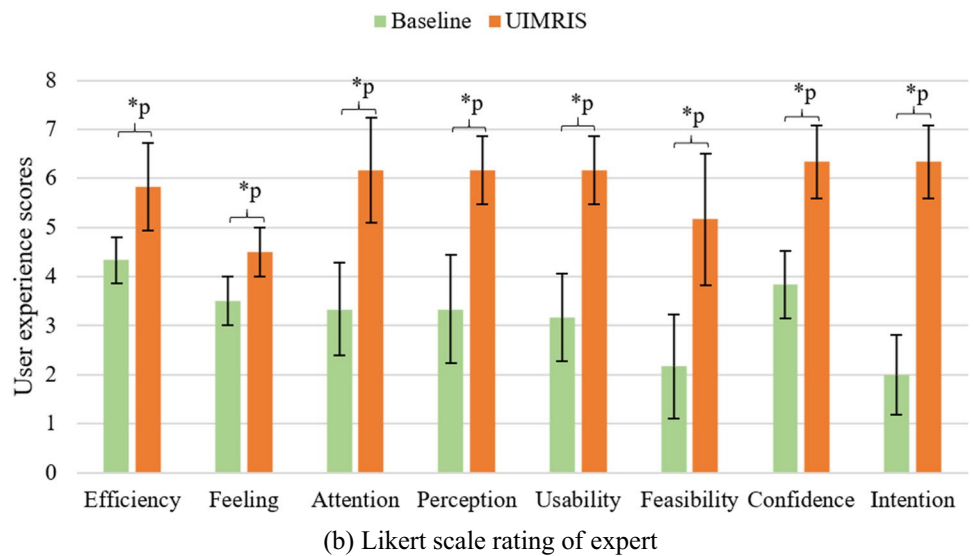
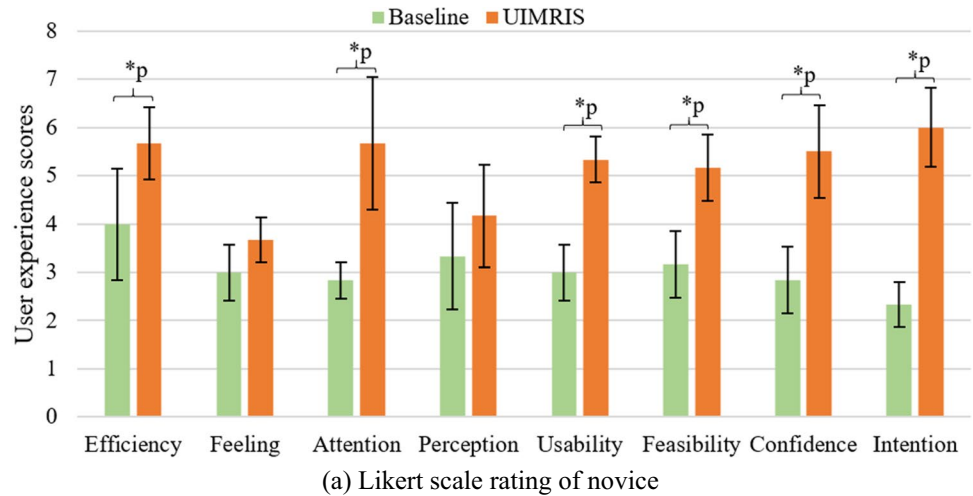
For novice (as shown in Fig. 14a), there were significant differences in terms of efficiency (Q1: $U = 31$, $p < 0.05$), attention (Q3: $U = 33.5$, $p = 0.009 < 0.01$), usability (Q5: $U = 36.0$, $p = 0.002 < 0.05$), feasibility (Q6: $U = 35.0$, $p = 0.004 < 0.05$), confidence (Q7: $U = 35.5$, $p = 0.002 < 0.05$), and intention (Q8: $U = 36.0$, $p = 0.002 < 0.05$). No significant differences between the two information visualization responses methods were observed for the other two factors (i.e., feeling (Q2: $U = 28$, $p = 0.132 > 0.05$), perception (Q4: $U = 24.5$, $p = 0.310 > 0.05$)).

For expert (as shown in Fig. 14b), there were significant differences in terms of efficiency (Q1: $U = 32$, $p < 0.05$), feeling (Q2: $U = 31.5$, $p < 0.05$), attention (Q3: $U = 34.5$, $p < 0.05$), perception (Q4: $U = 35.5$, $p < 0.05$), usability (Q5: $U = 36.0$, $p < 0.05$), feasibility (Q6: $U = 34.0$, $p < 0.05$), confidence (Q7: $U = 35.5$, $p < 0.05$), and intention (Q8: $U = 36.0$, $p < 0.05$).

4.6.5 Number of interactions

The number of interactions is an important indicator to evaluate the adaptive ability of a system [25]. Therefore, a

Fig. 14 User experience results (mean ± SD) for the two visual response methods reported by novice and expert, **p* indicates a significant difference between two different interfaces



comparison of the number of interactions between the two information visualization response methods was conducted. A two-way between-groups analysis of variance revealed the main effect of information visualization response methods on number of interactions was significant ($F(1,23)=168.438, p<0.001$). According to the single-variable test, the average number of interactions to complete the assembly task using UIAVS ($M=19.083$) interface was significantly shorter than that of Baseline ($M=30.333, p<0.001$), which was 37.1% shorter. There was a statistically main effect for level of expertise ($F(1,23)=365.998, p<0.001$). The single-variable test revealed that the number of interactions for the expert to complete the assembly task ($M=16.417$) was significantly shorter than that of the novice ($M=33.000, p<0.001$). The interaction effect between information visualization response methods and level of expertise was statistically significant ($F(1,23)=14.057, p<0.05$). Simple effect analysis showed that for novice, the number of interactions using UIAVS to complete assembly tasks ($M=29$) was 21.6% shorter than that

of Baseline [$M=37, p<0.001$]. In addition, for expert, the number of interactions used to complete the assembly task using UIAVS ($M=23.667$) was significantly shorter than Baseline ($M=9.167, p<0.001$), which was 61.3% shorter.

5 Discussion

In this section, the results of this user study were further discussed and analyzed. The authenticity of the five hypotheses was tested according to the experimental results in performance time, error evaluation, NASA-TLX, Likert scale, and number of interactions.

(1) Task performance

The performance time of two information visualization response methods in AR assembly in two groups of people with task familiarity (novice and expert) was tested to prove

hypothesis 1. First, according to Sect. 4.6.1, the average time to complete the assembly task using UIAVS interface is less than that of Baseline interface, which proves that UIAVS interface is more efficient (see Table 3). Feedback analysis from Question 1 also supports this view (see Fig. 14). According to Q1 and the time data of completing the task, finding effective and important information in the physical task is directly related to the user's operation response time, because the user needs to find the necessary assembly process information to know the precautions to correctly complete the assembly task. Therefore, it can be inferred that the greater the difficulty of information cognition, the longer it takes for users to find effective and important information, and the longer it takes to complete the task. Second, the average time to complete the assembly task is shorter for the expert than for the novice (see Table 3). This is consistent with the description in [63]. Third, Table 3 shows that under the premise of the same task familiarity level, the change of operation time reflects the change of user's information demand. After paying attention to the most important information, novices will still expand the LODs level to find other information they want. Therefore, it is only a little more efficient for novices to use UIAVS interface to complete assembly tasks than Baseline interface (see Table 3, Sect. 4.6.1). Experts have a higher cognitive level and pay more attention to the information they need. They use UIAVS interface to complete assembly tasks more efficiently than Baseline interface (Table 3 and Sect. 4.6.1). The feedback results of Q4 also support this view (see Fig. 14 and Sect. 4.6.4). "I know the assembly process, but I don't remember the parameters specified for assembly. I can easily find the parameters I want with this program. I just need to pay attention to these during assembly" said an expert who participated in UIAVS. A reasonable explanation for these different results is that UIAVS introduces a hierarchical mechanism of information importance, which speeds up the efficiency of users acquiring important and necessary information, so they can perform more efficiently. Therefore, Hypothesis 1 is finally accepted.

In order to prove Hypothesis 2, the correct assembly position of engine was verified by checking whether the parts meet the specified requirements. According to Table 3 and Sect. 4.6.2, there was no significant difference between the UIAVS interface and the Baseline interface, neither for novice nor expert. Essentially, users pay attention to the key information of parts assembly and form their own mental representations in memory through these information [64], thereby affecting the operation of assembly. In addition, it should be noted that the amount of assembly errors using the UIAVS interface was slightly lower than that of the Baseline interface, although there was no statistical difference between the two interfaces. This shows that the UIAVS interface has a certain role for users, which is consistent

with the feedback results of Q3 and Q5 (see Fig. 14 and Sect. 4.6.4). Through the interviews with participants and the analysis of assembly error parts, it can be inferred that the adaptive visualization of LODs of key information may have an impact on the psychological representation of users, which needs further research to link the visual characteristics of key information with the degree of user distraction reduction. Therefore, Hypothesis 2 is rejected.

(2) Information cognition

In order to prove Hypothesis 3, UIAVS adopts the hierarchical display of important information and the adaptive display of visualization, that is, by recognizing the user's intention to provide the user with necessary and required information in the engine assembly, so as to reduce the amount of information in the AR interface. It can be seen from Table 3 and Sect. 4.6.3 that the UIAVS interface can effectively reduce the cognitive load of users compared with the Baseline interface, both for novice and expert. In fact, the experiment proves the conclusion in the Ref [65], that is, reducing the interference of information in the interface can reduce the cognitive load of users. UIAVS improves the usability of visual information through user intention recognition (see Fig. 14), information processing (see Fig. 14), and reducing the time users spend searching for information (see Table 3). This enables the user to correctly complete the assembly task while reducing the amount of information, which is consistent with the feedback results of Q5 and Q7 (see Fig. 14 and Sect. 4.6.4). This might be attributed to the fact that UIAVS reduces the amount of information while ensuring the existence of necessary information, and gives users an interface for free interaction to obtain information, so that users can confidently complete assembly tasks. So Hypothesis 3 should be accepted.

(3) User experience

According to Fig. 14 and Sect. 4.6.4, the two interfaces have significant effects in Q1 (efficiency), Q3 (attention), Q5 (usability), Q6 (feasibility), Q7 (confidence), and Q8 (intention) for both novice and expert. UIAVS interface has better user experience than Baseline interface in these aspects. In other respects, there are some differences between novices and experts. For Q2 (feeling), there was no statistical difference among novices, which means that UIAVS interface has no obvious advantage in Q2 (feeling) over Baseline interface (see Fig. 14 and Sect. 4.6.4). Through interviews with participants, we found that the main reason for novices is the novelty effect; they always tend to expand LODs to understand all the information to ensure that they do not miss some information. For Q4 (perception), there was no statistical difference between the two interfaces among novices, but there was a significant difference among experts.

“In fact, what I am interested in is the data information. I know how to operate it, but the data is too difficult to remember. When I assemble parts, the process data information will always pop up. This is really great.” An expert who participated in UIAVS experiment said so. For Q4 (perception), the user experience of the expert users who participated in the UIAVS interface experiment was significantly better than that of the Baseline interface. However, for Q4, the user experience of novices participating in UIAVS interface experiment had no advantage over that of baseline interface. Based on interviews with participants and data analysis, it can be speculated that the reason for this phenomenon is largely because novices do not know the main information they want, so they will still expand the LODs level to obtain comprehensive information. In addition, the experts know the information they want so they can easily find it. Therefore, Hypothesis 4 is selectively accepted.

(4) Interaction behavior

As can be seen from Table 3 and Sect. 4.6.5, there is a significant difference in the number of interactions between the two interfaces. Because UIAVS interface can adaptively adjust the visualization form according to the user’s intention, users do not often need to change the distribution of information in AR interface through manual interaction. This effectively reduces the number of user interactions. The feedback from Q3, Q4, and Q8 also proves this view (see Fig. 14 and Sect. 4.6.4). Changes in the number of interactions also reflect changes in user information needs (see Sect. 4.6.5). Experts have higher cognitive levels, so they focus on less information, and the number of interactions using the UIAVS interface is significantly less than the Baseline interface (see Table 3 and Sect. 4.6.5). Relatively speaking, novices have lower cognitive level and pay more attention to information, so they have more interactions. Therefore, the number of interactions using UIAVS is slightly lower than that of Baseline (see Table 3 and Sect. 4.6.5). A reasonable explanation for these different results is that UIAVS introduces an adaptive visualization adjustment function and a hierarchical mechanism of information importance, which enables the system to automatically adjust changes according to user intentions, thereby effectively reducing the number of interactions. Therefore, Hypothesis 5 is finally accepted.

6 Limitations and future works

(1) Simple physical assembly tasks

The experiments simulated the assembly of a small engine on the assembly table. The assembly tasks are

simpler than those of large-scale parts in real-world scenarios. It does not involve the assembly of curved space parts and complex process drawing requirements. However, it does have some key elements in assembly, such as assembly position requirements, assembly precautions, and assembly process requirements. This may have some limitations on the applicability of the experimental results. Therefore, in future work, it is needed to verify the results of the experiments in more scenarios, such as high-precision assembly and complex assembly environments.

(2) Smooth transition settings

Through interviews with participants, it was found that adaptive adjustment of visualization forms in the UIAVS interface can be confusing to users. Although the system has added a progress bar-like indication mechanism to the change of the visual form. In addition, this experiment cannot prove that Q2 may also be related to unreasonable changes in visual form (see Fig. 14 and Sect. 5). Therefore, in the future, we can add a time smooth transition setting when the visual form changes adaptively on the existing basis, so as to minimize the trouble caused to users by the sudden change of the visual form.

7 Conclusion

This paper is the first to propose a method for recognizing user intention to adaptively adjust information in AR manual assembly. This paper proves that the system (UIAVS) has higher user performance than the traditional AR assembly instruction method, especially for expert. The purpose of this study is to recognize the user’s intention based on the user’s gaze, spatial location, and the current assembly task, and to adaptively adjust the display form of virtual information to highlight the information needed by the user. UIAVS was developed by establishing an information hierarchy mechanism, and using a logic rule-based method to analyze the user’s intention and to adaptively provide the corresponding level of visualization information according to the intention. An experimental case was designed to imitate the actual engine assembly. To test the effect of the experiment, 24 participants were randomly assigned to different task familiarity levels (novice and expert) and different information visualization response methods (UIAVS and Baseline). The experimental results were analyzed in terms of performance time, error evaluation, cognitive load, user experience, and number of interactions. All hypotheses except Hypothesis 2 were accepted. Therefore, UIAVS is helpful for improving the cognitive efficiency of users.

This paper is just a preliminary exploration of recognizing users' intention to adjust the form of information visualization. In fact, the method proposed in this paper can be used not only in AR manual assembly, but also in other fields such as AR maintenance and AR education. In addition, this paper is user-centered to reduce the cognitive burden of the assembly site personnel as much as possible. However, process designers are required to rank important information levels, which increases the burden of process designers to a certain extent. Therefore, future research work will further investigate the automatic information hierarchy ranking algorithm to reduce the burden on process designers and make up for the deficiencies in this research.

Acknowledgements We would like to appreciate Professor Weiping He, Xiaoliang Bai, and Shuxia Wang for their science leadership, and the constructive opinions of Dr. Xiangyu Zhang significantly improved the paper. We also would like to appreciate Peng Wang for donating the engine model used in this research. In addition, we would like to thank Liwei Liu and Qing Yu for helping experiment with data collection and preparing some supplementary materials.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Y. Y., X. Z., L. L., and Q. Y. The first draft of the manuscript was written by Y. Y. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was partly supported by National Key R&D Program of China (Grant numbers 2019YFB1703800, 2021YFB1714900, 2021YFB1716200, and 2020YFB1712503), the Programme of Introducing Talents of Discipline to Universities (111 Project), China (Grant number B13044), the Fundamental Research Funds for the Central Universities, NPU (Grant number 3102020gxb003), and General Project of Chongqing Natural Science Foundation (Grant number cstb2022nscq-msx1153).

Declarations

Competing interests The authors declare no competing interests.

References

- Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. *IEICE Trans Inf Syst* 77(12):1321–1329
- Yun H, Jun MBG (2022) Immersive and interactive cyber-physical system (I2CPS) and virtual reality interface for human involved robotic manufacturing. *J Manuf Syst* 62:234–248. <https://doi.org/10.1016/j.jmsy.2021.11.018>
- Chu CH, Ko CH (2021) An experimental study on augmented reality assisted manual assembly with occluded components. *J Manuf Syst* 61:685–695. <https://doi.org/10.1016/j.jmsy.2021.04.003>
- Chu CH, Wang L, Liu S, Zhang Y, Menozzi M (2021) Augmented reality in smart manufacturing: enabling collaboration between humans and artificial intelligence. *J Manuf Syst* 61:658–659. <https://doi.org/10.1016/j.jmsy.2021.05.006>
- Dalle Mura M, Dini G (2021) Augmented reality in assembly systems: state of the art and future perspectives. *Smart Technologies for Precision Assembly: 9th IFIP WG 5.5 International Precision Assembly Seminar, IPAS 2020, Virtual Event, December 14–15, 2020, Revised Selected Papers 9*. Springer International Publishing, pp 3–22. https://doi.org/10.1007/978-3-030-72632-4_1
- Wiedenmaier S, Oehme O, Schmidt L, Luczak H (2003) Augmented reality (AR) for assembly processes design and experimental evaluation. *Int J Hum-Comput Interact* 16(3):497–514. https://doi.org/10.1207/S15327590IJHC1603_7
- Goher K, Shehab E, Al-Ashaab A (2021) Model-based definition and enterprise: state-of-the-art and future trends. *Proc Inst Mech Eng, Part B: J Eng Manuf* 235(14):2288–2299. <https://doi.org/10.1177/0954405420971087>
- Quintana V, Rivest PR, Venne F, Kheddouci F (2010) Will model-based definition replace engineering drawings throughout the product lifecycle? A global perspective from aerospace industry. *Comput Ind* 61(5):497–508. <https://doi.org/10.1016/j.compind.2010.01.005>
- Wang Z, Bai X, Zhang S, Billingham M, He W, Wang P, Lan W, Min H, Chen Y (2022) A comprehensive review of augmented reality-based instruction in manual assembly, training and repair. *Robot Comput-Integr Manuf* 78:102407. <https://doi.org/10.1016/j.rcim.2022.102407>
- Kim YM, Rhiu I, Yun MH (2020) A systematic review of a virtual reality system from the perspective of user experience. *Int J Hum-Comput Interact* 36(10):893–910. <https://doi.org/10.1080/10447318.2019.1699746>
- Michalos G, Karagiannis P, Makris S, Tokçalar Ö, Chryssolouris G (2016) Augmented reality (AR) applications for supporting human-robot interactive cooperation. *Procedia CIRP* 41:370–375. <https://doi.org/10.1016/j.procir.2015.12.005>
- Lai ZH, Tao W, Leu MC, Yin Z (2020) Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. *J Manuf Syst* 55:69–81. <https://doi.org/10.1016/j.jmsy.2020.02.010>
- Eswaran M, Gulivindala A K, Inkulu A K, Raju Bahubalendruni M V A (2023) Augmented reality-based guidance in product assembly and maintenance/repair perspective:: a state of the art review on challenges and opportunities. *Expert Syst Appl* 213. <https://doi.org/10.1016/j.eswa.2022.118983>
- Neumann U, Majoros A (1998) Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No. 98CB36180)*. IEEE, pp 4–11. <https://doi.org/10.1109/VRAIS.1998.658416>
- Sausman J, Samoylov A, Regli SH, Hopps M (2012) Effect of eye and body movement on augmented reality in the manufacturing domain. *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, pp 315–316. <https://doi.org/10.1109/ISMAR.2012.6402591>
- Tang A, Owen C, Biocca F, Mou W (2003) Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, pp 73–80. <https://doi.org/10.1145/642611.642626>
- Kolla S S V K, Sanchez A, Plapper P (2021) Comparing effectiveness of paper based and augmented reality instructions for manual assembly and training tasks. In *Proceedings of the Conference on Learning Factories (CLF)*, Graz, Austria. Available at <https://doi.org/10.2139/ssrn.3859970>
- Westerfield G, Mitrovic A, Billingham M (2015) Intelligent augmented reality training for motherboard assembly. *Int J Artif Intell Educ* 25:157–172. <https://doi.org/10.1007/s40593-014-0032-x>
- Blattgerste J, Strenge B, Renner P, Pfeiffer T, Essig K (2017) Comparing conventional and augmented reality instructions for manual assembly tasks. In *Proceedings of the 10th international conference on pervasive technologies related to assistive environments*. ACM, pp 75–82. <https://doi.org/10.1145/3056540.3056547>

20. Salonen T, Sääski J, Hakkarainen M, Kannetis T, Perakakis M, Siltanen S, Potamianos A, Korkalo O, Woodward C (2007) Demonstration of assembly work using augmented reality. In Proceedings of the 6th ACM international conference on image and video retrieval. ACM, pp 120–123. <https://doi.org/10.1145/1282280.1282301>
21. Lee H, Bang S, Woo W (2020) Effects of background complexity and viewing distance on an ar visual search task. 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE, pp 189–194. <https://doi.org/10.1109/ISMAR-Adjunct51615.2020.00057>
22. Kruijff E, Swan JE, Feiner S (2010) Perceptual issues in augmented reality revisited. 2010 IEEE International Symposium on Mixed and Augmented Reality. London, IEEE, pp 3–12. <https://doi.org/10.1109/ISMAR.2010.5643530>
23. Radkowski R, Herrema J, Oliver J (2015) Augmented reality-based manual assembly support with visual features for different degrees of difficulty. *Int J Hum-Comput Interact* 31(5):337–349. <https://doi.org/10.1080/10447318.2014.994194>
24. Li W, Wang J, Jiao S, Wang M, Li S (2019) Research on the visual elements of augmented reality assembly processes. *Virtual Reality Intell Hardw* 1(6):622–634. <https://doi.org/10.1016/j.vrih.2019.09.006>
25. Lindlbauer D, Feit A M, Hilliges O (2019) Context-aware online adaptation of mixed reality interfaces. In Proceedings of the 32nd annual ACM symposium on user interface software and technology. ACM, pp 147–160. <https://doi.org/10.1145/3332165.3347945>
26. Yang Z, Shi J, Jiang W, Sui Y, Wu Y, Ma S, Kang C, Li H (2019) Influences of augmented reality assistance on performance and cognitive loads in different stages of assembly task. *Front Psychol* 10:1703. <https://doi.org/10.3389/fpsyg.2019.01703>
27. Geng J, Song X, Pan Y, Tang J, Liu Y, Zhao D, Ma Y (2020) A systematic design method of adaptive augmented reality work instruction for complex industrial operations. *Comput Ind* 119:103229. <https://doi.org/10.1016/j.compind.2020.103229>
28. Kaghat FZ, Azough A, Fakhour M, Meknassi M (2020) A new audio augmented reality interaction and adaptation model for museum visits. *Comput Electr Eng* 84:106606. <https://doi.org/10.1016/j.compeleceng.2020.106606>
29. Pfeuffer K, Abdrabou Y, Esteves A, Rivu R, Abdelrahman Y, Meitner S, Saadi A, Alt F (2021) ARtention: a design space for gaze-adaptive user interfaces in augmented reality. *Comput Graph* 95:1–12. <https://doi.org/10.1016/j.cag.2021.01.001>
30. Gardony AL, Lindeman RW, Brunyé TT (2020) Eye-tracking for human-centered mixed reality: promises and challenges. *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR)*, vol 11310. SPIE, pp 230–247. <https://doi.org/10.1117/12.2542699>
31. Xia L, Lu J, Zhang Z, Chen R, Wang S, Zhang H (2019) Development and application of parts assembly guidance system based on augmented reality. 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), vol 1. IEEE, pp 1325–1330. <https://doi.org/10.1109/IAEAC47372.2019.8997861>
32. Dalle Mura M, Dini G (2021) An augmented reality approach for supporting panel alignment in car body assembly. *J Manuf Syst* 59:251–260. <https://doi.org/10.1016/j.jmsy.2021.03.004>
33. Thomas P C, David W M (1992) Augmented reality: an application of heads-up display technology to manual manufacturing processes. In Hawaii international conference on system sciences, Kauai, HI, USA, ACM SIGCHI Bulletin. IEEE 2:659–669. <https://doi.org/10.1109/HICSS.1992.183317>
34. Wang Z, Bai X, Zhang S, He W, Zhang X, Zhang L, Wang P, Han D, Yan Y (2020) Information-level AR instruction: a novel assembly guidance information representation assisting user cognition. *Int J Adv Manuf Technol* 106(1):603–626. <https://doi.org/10.1007/s00170-019-04538-9>
35. Raptis GE, Fidas C, Avouris N (2018) Effects of mixed-reality on players' behaviour and immersion in a cultural tourism game: a cognitive processing perspective. *Int J Hum Comput Stud* 114:69–79. <https://doi.org/10.1016/j.ijhcs.2018.02.003>
36. Khamis M, Alt F, Bulling A (2016) Challenges and design space of gaze-enabled public displays. In Proceedings of the 2016 ACM International joint conference on pervasive and ubiquitous computing: Adjunct ACM, pp 1736–1745. <https://doi.org/10.1145/2968219.2968342>
37. Ichihashi K, Fujinami K (2019) Estimating visibility of annotations for view management in spatial augmented reality based on machine-learning techniques. *Sensors* 19(4):939. <https://doi.org/10.3390/s19040939>
38. Julier S, Lanzagorta M, Baillet Y, Rosenblum L, Feiner S, Hollerer T, Sestito S (2000) Information filtering for mobile augmented reality. Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000). IEEE, pp 3–11. <https://doi.org/10.1109/ISAR.2000.880917>
39. Feiner S, MacIntyre B, Seligmann D (1993) Knowledge-based augmented reality. *Commun ACM* 36(7):53–62. <https://doi.org/10.1145/159544.159587>
40. Zhu J, Ong SK, Nee AY (2015) A context-aware augmented reality assisted maintenance system. *Int J Comput Integr Manuf* 28(2):213–225. <https://doi.org/10.1080/0951192X.2013.874589>
41. Feiner S, MacIntyre B, Haupt M, Solomon E (1993) Windows on the world: 2D windows for 3D augmented reality. In Proceedings of the 6th annual ACM symposium on User interface software and technology. ACM, pp 145–155. <https://doi.org/10.1145/168642.168657>
42. DiVerdi S, Hollerer T, Schreyer R (2004) Level of detail interfaces. Third IEEE and ACM International Symposium on Mixed and Augmented Reality. IEEE, pp 300–301. <https://doi.org/10.1109/ISMAR.2004.38>
43. Tatzgern M, Orso V, Kalkofen D, Jacucci G, Gamberini L, Schmalstieg D (2016) Adaptive information density for augmented reality displays. 2016 IEEE Virtual Reality (VR). IEEE, pp 83–92. <https://doi.org/10.1109/VR.2016.7504691>
44. Ghouaïel N, Cieutat J M, Jessel J P (2014) Adaptive augmented reality: plasticity of augmentations. In Proceedings of the 2014 virtual reality international conference. ACM, pp 1–4. <https://doi.org/10.1145/2617841.2620695>
45. Tecchia F, Alem L, Huang W (2012) 3D helping hands: a gesture based MR system for remote collaboration. In Proceedings of the 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry. ACM, pp 323–328. <https://doi.org/10.1145/2407516.2407590>
46. McNamara A, Boyd K, George J, Jones W, Oh S, Suther A (2019) Information placement in virtual reality. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, pp 1765–1769. <https://doi.org/10.1109/VR.2019.8797891>
47. Rathnayake D, de Silva A, Puwakandawa D, Meegahapola L, Misra A, Perera I (2020) Jointly optimizing sensing pipelines for multimodal mixed reality interaction. 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS). IEEE, pp 309–317. <https://doi.org/10.1109/MASS50613.2020.00046>
48. Seipel P, Stock A, Santhanam S, Baranowski A, Hochgeschwender N, Schreiber A (2019) Speak to your software visualization—exploring component-based software architectures in augmented reality with a conversational interface. 2019 Working Conference on Software Visualization (VISOFT). IEEE, pp 78–82. <https://doi.org/10.1109/VISOFT.2019.00017>
49. Pfeuffer K, Mayer B, Mardanbegi D, Gellersen H (2017) Gaze+pinch interaction in virtual reality. In Proceedings of the 5th symposium on spatial user interaction. ACM, pp 99–108. <https://doi.org/10.1145/3131277.3132180>

50. Wang Z, Bai X, Zhang S, He W, Wang P, Zhang X, Yan Y (2020) SHARIdeas: a visual representation of intention sharing between designer and executor supporting AR assembly. In SIGGRAPH Asia 2020 Posters. ACM, pp 1–2. <https://doi.org/10.1145/3415264.3431858>
51. Piumsomboon T, Dey A, Ens B, Lee G, Billingham M (2017) [POSTER] CoVAR: mixed-platform remote collaborative augmented and virtual realities system with shared collaboration cues. 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct). IEEE, pp 218–219. <https://doi.org/10.1109/ISMAR-Adjunct.2017.72>
52. Sonntag D, Albuquerque G, Magnor M, Bodensiek O (2019) Hybrid learning environments by data-driven augmented reality. *Procedia Manuf* 31:32–37. <https://doi.org/10.1016/j.promfg.2019.03.006>
53. Aggarwal JK, Ryoo MS (2011) Human activity analysis: a review. *Acm Comput Surv (Csur)* 43(3):1–43. <https://doi.org/10.1145/1922649.1922653>
54. Bernardos AM, Gómez D, Casar JR (2016) A comparison of head pose and deictic pointing interaction methods for smart environments. *Int J Hum-Comput Interact* 32(4):325–351. <https://doi.org/10.1080/10447318.2016.1142054>
55. Jacob R J (1990) What you look at is what you get: eye movement-based interaction techniques. In Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp 11–18. <https://doi.org/10.1145/97243.97246>
56. Zhang Y, Ding K, Hui J, Lv J, Zhou X, Zheng P (2022) Human-object integrated assembly intention recognition for context-aware human-robot collaborative assembly. *Adv Eng Inform* 54:101792. <https://doi.org/10.1016/j.aei.2022.101792>
57. Hart S G, Staveland L E (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In *Advances in psychology. Human mental workload*, vol. 52 pp 139–183. North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
58. Hinderks A, Schrepp M, Mayo FJD, Escalona MJ, Thomaschewski J (2019) Developing a UX KPI based on the user experience questionnaire. *Comput Stan Interfaces* 65:38–44. <https://doi.org/10.1016/j.csi.2019.01.007>
59. Schrepp M (2015) User experience questionnaire handbook: all you need to know to apply the ueq successfully in your projects. <https://www.ueq-online.org/>. Accessed 22 May 2023
60. Chanquoy L, Tricot A, Sweller J (2007) *La charge cognitive: théorie et applications [the cognitive load: theories and applications]*. Paris: Armand Colin
61. Megalakaki O, Tijus C, Baiche R, Poitrenaud S (2012) The effect of semantics on problem solving is to reduce relational complexity. *Think Reason* 18(2):159–182. <https://doi.org/10.1080/13546783.2012.663101>
62. Tcha-Tokey K, Christmann O, Loup-Escande E, Richir S (2016) Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *Int J Virtual Real* 16(1):33–48. <https://doi.org/10.20870/IJVR.2016.16.1.2880>
63. Abernethy B, Russell DG (1987) Expert-novice differences in an applied selective attention task. *J Sport Exerc Psychol* 9(4):326–345. <https://doi.org/10.1123/jsp.9.4.326>
64. Wang Z, Bai X, Zhang S, Billingham M, He W, Wang Y, Han D, Chen G, Li J (2021) The role of user-centered AR instruction in improving novice spatial cognition in a high-precision procedural task. *Adv Eng Inform* 47:101250. <https://doi.org/10.1016/j.aei.2021.101250>
65. Reis HM, Borges SS, Durelli VH, Moro LFDS, Brandao AA, Barbosa EF, Brandão LO, Isotani S, Jaques PA, Bittencourt II (2012) Towards reducing cognitive load and enhancing usability through a reduced graphical user interface for a dynamic geometry system: an experimental study. 2012 IEEE International Symposium on Multimedia. IEEE, pp 445–450. <https://doi.org/10.1109/ISM.2012.91>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.