Virtual Reality
https://doi.org/10.1007/s10055-022-00742-3

ORIGINAL ORIGINAL

X-Board: an egocentric adaptive AR assistant for perception in indoor environments

Zhenning Zhang[1] · Zhigeng Pan[2] · Weiqing Li[1] · Zhiyong Su[3]

Received: 29 April 2022 / Accepted: 15 December 2022

Abstract
Augmented reality (AR) has the potential to become an effective assistive technology for emergencies in the future. However, raw AR content can confuse users' visual perception and occlude information in the physical world. In this research, we propose X-Board, an X-ray visualization-based AR assistant for perception in indoor environments. In accordance with its design principles, X-Board provides visual–spatial cues by means of a grid mesh corresponding to the occluding surface in front of the target object. Meanwhile, the X-Board interacts with the physical world in real time, improving the coherence between the virtual and real worlds. To ensure the appropriate allocation of the user's visual resources, the user's visual intention is recognized based on gaze data to realize an adaptive display feature. The results of the user evaluation show that X-Board can effectively improve the accuracy and speed of the perception and reduce the cognitive load on users; thus, the usability of X-Board is confirmed. With X-Board, users could effectively perceive the spatial positions of their comrades in an indoor occluded environment in our simulated perception scenario.

Keywords Augmented reality · Perception · Human–Computer interaction · X-ray visualization-based · Adaptive display · Visual resources

1 Introduction

Situational awareness (SA) is defined as "the perception of environmental elements and events with respect to time or space, the comprehension of their meaning, and the projection of their future status" (Endsley 1995). Considering that the purpose of augmented reality (AR) is to mix the physical world and virtual world by overlaying virtual information on a person's real environment, we believe that appropriate AR content can improve user ability of perception (i.e. the first level of SA), so as to improve user performance in SA tasks. In fact, AR has already been studied and applied for a variety of primitive SA tasks (i.e. location and navigation tasks), such as marine navigation (Hertel and Steinicke 2021), firefighting (Dominic and Robb 2020), and military (Nilsson et al. 2009; Livingston et al. 2011) applications. AR has the ability to extend the sources of visual information by overlaying previously invisible information, extending visual perception, and enhancing human cognition; thus, it has considerable potential in SA tasks.

In particular, an AR visualization method can improve the perception ability in emergencies. To this end, Osmers et al. compared three AR-based mechanisms (map, X-ray, and compass) (Osmers and Prilla 2020). Their results showed that the map and X-ray mechanisms performed best for awareness and orientation. However, we believe that raw AR visualization mechanisms still need improvement to efficiently and naturally display assistance information. Considering that reliance on a map tends to lead to more context switching between the virtual and real worlds and a greater cognitive load than the X-ray vision method (Dey et al. 2011), and that indoor digital maps are often difficult to obtain for emergency tasks, we focus our work on

✉ Zhenning Zhang
zhenningzhang@njust.edu.cn

Zhigeng Pan
zgpan@hznu.edu.cn

Weiqing Li
li_weiqing@njust.edu.cn

1 School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

2 College of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing, China

3 School of Automation, Nanjing University of Science and Technology, Nanjing, China

Published online: 26 December 2022

X-ray-based AR visualization technologies to improve users' perception.

Despite the potential advantages, X-ray-based AR content overlaid on the real environment will also present new challenges to human visual perception. On the one hand, the AR content must be able to accurately represent spatial relationships such as occlusion layers (Livingston et al. 2003) or depth cues (Vaziri et al. 2021); on the other hand, users with AR head-worn displays (HWDs) will need to balance the allocation of their visual resources (i.e. attention) between virtual and real elements to avoid missing key information in the physical environment. These two factors are particularly important for an AR visualization method for our perception task (i.e. a simulated counterterrorism scenario). Occluded visualization and depth estimation are important research directions of AR visualization. However, to provide effective assistance for indoor perception tasks, an AR visualization method also needs to support the feature of adaptive display based on the user's intention to avoid occluding important information in the real environment.

In this research, we propose X-Board, an X-ray visualization-based AR assistant for perception in indoor environments. Combined with the requirements of the perception task, X-Board is designed to provide visual–spatial cues by means of a grid mesh corresponding to the occluding surface in front of the target object based on design principles suited to the task of interest. Meanwhile, X-Board interacts with the physical world in real time, improving the coherence between the virtual and real worlds. To ensure appropriate allocation of the user's visual resources, a long short-term memory (LSTM) neural network is employed to recognize the user's visual intention (i.e. whether the user is focused on the "nearby real environment" or "distant virtual objects") based on gaze sequence data and to control the AR assistive display accordingly. We have evaluated our X-Board prototype under three conditions (Fig. 1a–c) in the perception task scenario. The results of the user evaluation and user feedback confirm the effectiveness and usability of X-Board.

Our main contributions are summarized as follows:

- We propose X-Board, an X-ray visualization-based AR assistant for perception tasks in indoor environments. X-Board provides visual–spatial cues and coherence to improve the visual perception, and adaptive display by the user's gaze data to ensure the appropriate allocation of the visual resources.
- A user study has been conducted to evaluate the usability and effectiveness of X-Board. The results of the user evaluation show that X-Board can effectively improve the accuracy and speed of perception and reduce the cognitive load on users in perception tasks.

The remainder of this paper is structured as follows: Sect. 2 reviews related work. In Sect. 3, we analyse the

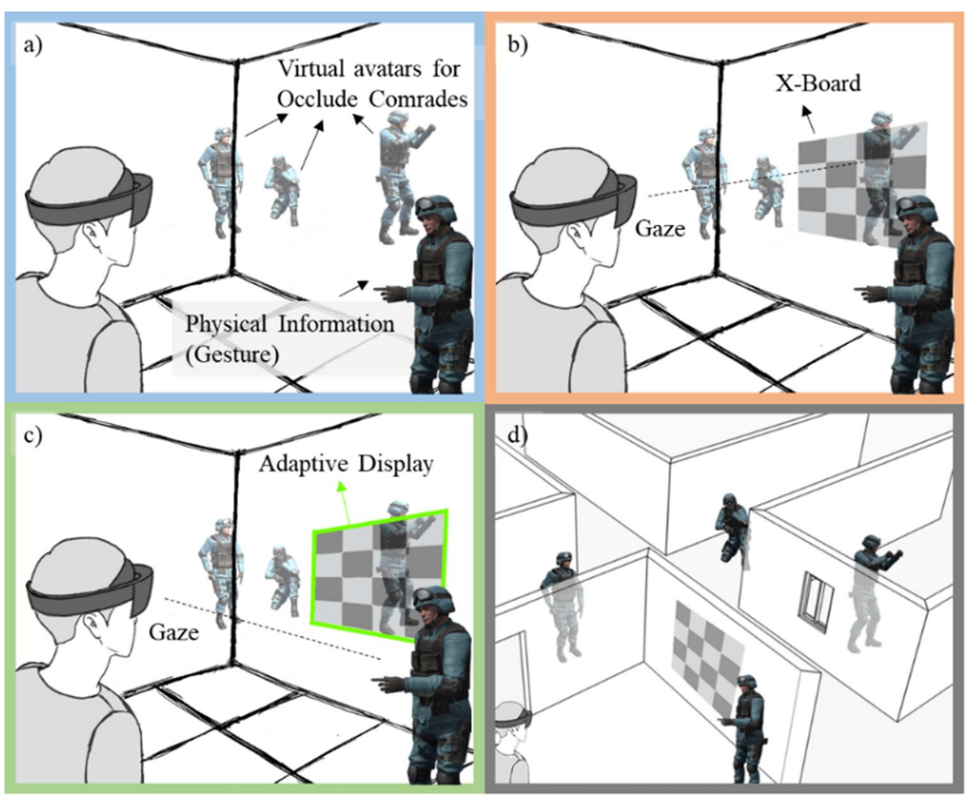


Fig. 1 The X-ray visualization-based adaptive AR assistant proposed in this paper: **a** virtual avatars for occluded comrades overlaid on the user's view without assistance; **b** visualization with X-Board, which provides an occlusion layer and depth cues by means of spatial interaction and a grid mesh; **c** to balance the allocation of visual resources, adaptive display of X-Board on the basis of the user's real-time eye gaze data; **d** aerial view of our simulated counterterrorism scenario

requirements of perception tasks in indoor environments and investigate corresponding design principles for X-Board. Section 4 presents our implementation method, in which X-Board can interact with the physical world, switching targets based on the direction of the user's head gaze and providing an adaptive display function based on the user's eye gaze data in real time. We report experiments and corresponding results in Sects. 5 and 6 and discuss them in Sect. 7. In Sect. 8, we report the limitations of our work and prospects for future work. The paper is concluded in Sect. 9.

2 Related work

2.1 Visual perception in AR

For a perception task in an indoor environment, it is necessary to provide the user with essential occlusion and depth cues to allow the user to understand the relationships between the virtual information and the physical world.

2.1.1 Occlusion in AR

AR extends the sources of visual perception, allowing users to gain "supernatural" powers such as X-ray vision, which can reduce the cognitive load of tasks such as manipulating occluded objects (Lilija et al. 2019) or controlling a drone's flight (Erat et al. 2018). Zollmann et al. (2010, 2014) demonstrated that image-based ghosting helps the user understand the subsurface locations of virtual objects. Bane et al. presented a set of interactive tools designed to give users virtual X-ray vision, visualizing the interiors of a small set of buildings (Bane and Hollerer 2004). Sandor et al. used visual saliency to provide users with richer context for occluded objects and obtained encouraging results (Sandor et al. 2010). Barnum et al. (2009) presented a method to create an illusion of seeing moving objects through occluding surfaces in a video. Gruenefeld et al. (2020) found that utilizing an additional aid (textual or grid-based) helped users estimate the distances to occluded objects more accurately. Tsuda et al. (2006) evaluated five visualization methods for X-ray vision. The results indicated that the optimal combination was to show a ground grid, overlay wireframe models of occluded objects, and provide a top-down view. Eren et al. (2013) proposed a multiview visualization technique that enables effortless interaction with subterranean data and attempts to maximize spatial perception while minimizing view clutter. Avery et al. (2009) presented multiple view modes to support new visualizations that provide depth cues and spatial awareness to users.

Direct occlusion visualization imposes a high cognitive load because the AR content is similar to an overlay in the user's field of view, violating the occlusion hierarchy of normal perception. In our research, X-Board is designed to improve the perception of occluded objects by enhancing the coherence between real and virtual objects.

2.1.2 Depth in AR

Estimation of the depth of virtual elements is a key issue in AR perception. Many previous works have investigated depth estimation under various conditions, such as at short and long distances (Hertel and Steinicke 2021; Liu et al. 2020; Swan et al. 2015; Gagnon et al. 2021), in indoor and outdoor environments (Dey et al. 2010; Livingston et al. 2009; Wither and Hollerer 2005), and in AR (Peillard et al. 2019) and virtual reality (VR) (Ping et al. 2019) modes, as well as other factors (Hertel and Steinicke 2021; Vaziri et al. 2021; Diaz et al. 2017) concerning the interaction of virtual objects with the real world. In our study, we focus on visualization technologies intended for use with an AR HWD, as they have the potential to improve the user's depth perception. Furmanski et al. found that additional occlusion and motion cues are necessary for relative depth judgement (Furmanski et al. 2002). Uratani et al. (2005) proposed depth visualization techniques in which the visual attributes of each annotation, such as the colour or line style of the frame, are dynamically modified to facilitate the perception of the annotations' locations and spatial relationships. Kytö et al. (2013) presented a visualization approach for improving relative depth judgements in AR. The results showed that the presence of auxiliary objects significantly reduced errors in depth judgement. Livingston et al. (2003) identified drawing style and opacity settings that enabled users to accurately interpret three layers of occluded objects, even in the absence of perspective constraints. In the present work, X-Board is assigned a translucent mesh with grid markings to assist users in depth estimation.

2.2 Spatial tasks with AR

AR has been adopted in many domain spatial tasks, such as piloting (Minaskan et al. 2021), health care (Pascale et al. 2019), maritime (Sharma et al. 2019) and military (Nilsson et al. 2009; Livingston et al. 2011) operations. We investigate two primitive tasks that are strongly related to AR: localization and navigation. Gruenefeld et al. (2019) compared a printed map to three different AR visualization strategies (in-view, out-of-view and combined). The results showed that in-view visualization reduces the error rates for object selection accuracy. Osmers et al. (2020) compared the effectiveness of three AR-based mechanisms (map, X-ray, and compass) when searching for and coordinating with objects. The results showed that the map and X-ray mechanisms performed best. Zhang et al. (2021) explored the impact of immersive augmented reality (IAR)-based systems on human wayfinding performance from the cognitive perspective and found that AR has the potential to reduce the human workload for cognitive tasks.

However, existing AR-based spatial research has given considerable attention to the content of the AR visualizations while ignoring the information of the dynamic physical world. Therefore, our goal is for our AR technology to better balance the user's visual attention between virtual and real elements, making it easier for the user to perceive important visual information.

2.3 Adaptive display in AR

In perception tasks, real elements are as important as virtual elements. If there are too many virtual elements, the cognitive load will be too high, and the user will become confused. In the Glanceable AR paradigm (Lu et al. 2020), secondary information is placed at the periphery of the visual field so that it will remain unobtrusive but can be accessed by a quick glance whenever needed. Gebhardt et al. (2019) proposed a reinforcement learning method to learn when to show or hide an object's label based on eye movement data. Sidenmark et al. (2019) proposed leveraging the synergetic movements of the eye and head and identified corresponding design principles for eye-and-head gaze interaction. Pfeuffer et al. (2021) presented AR tention, a design space for gaze interaction specifically tailored for in situ AR information interfaces. Lindlbauer et al. (2019) presented an optimization-based approach to automatically control when and where virtual labels are shown and how much they display. Alghofaili et al. developed a novel adaptive aid that maintains the effectiveness of traditional aids while providing designers and users with control of how often help is displayed (Gruenefeld et al. 2019). Inspired by the adaptive aids (Gruenefeld et al. 2019) provided in VR environments, a data-driven method is adopted here to adaptively display X-Board in order to realize an appropriate distribution of visual resources between virtual and real content.

3 Design of X-Board

We aim to propose an X-ray visualization-based AR assistant for perception tasks in complex indoor environments to help users accurately perceive the environment. An X-ray visualization-based AR visualization technique offers the capacity to see through walls, which has obvious advantages for indoor perception tasks. However, designing an X-ray visualization-based AR visualization technique requires considering not only how to improve the user's perception but also how to display the virtual information necessary for indoor tasks (e.g. firefighting or counterterrorism).

In this section, we introduce the requirements of perception tasks in indoor environments. In addition, oriented towards such perception tasks, we detail the design principles of X-Board.

3.1 The perception tasks in indoor environment

Research on indoor localization and navigation trails decades behind research on outdoor localization and navigation (Winter et al. 2019). Because of the higher accuracy required for indoor navigation and localization, combined with the fragmentation and compartmentalization of indoor environments, it is more difficult to complete indoor perception tasks.

In this work, we simulated an indoor counterterrorism scenario. In this scenario, occluded comrades are displayed as virtual avatars in AR. We mainly focus on the localization task in an occluded indoor environment. Our goal is to allow users to quickly and accurately perceive the locations of multiple targets (occluded comrades) using the assistive technology of X-Board. In the meantime, users should not sacrifice their visual attention to the real world.

3.2 Design principles

For localization and navigation tasks, reliance on a map tends to lead to more context switching between the virtual and real worlds and heavier cognitive load than an X-ray visualization-based mechanism (Dey et al. 2011). Moreover, indoor digital maps are often difficult to obtain. Thus, we wish to provide location information to the user through an egocentric X-ray view. To achieve this goal, we refer to the AR visualization taxonomy (Zollmann et al. 2020) and propose three design principles for our X-Board assistive technology:

DP1. Visual Spatial Cues: occlusion and depth cues should be provided to the user.

DP2. Visual Coherence: virtual content should be integrated into the physical world.

DP3. Adaptive Display: virtual content should be displayed adaptively to avoid occluding real-world information.

3.2.1 DP1. Visual spatial cues

Visual spatial cues include depth cues and occlusion cues, which help users better understand how abstract AR graphics are related to the space in which they are displayed. Livingston et al. found that representations based on wireframe outlines and decreasing opacity have advantages in conveying information on far-field occluded objects to the user (Livingston et al. 2003), and the results of Kytö et al. (2013) showed that the presence of auxiliary objects significantly reduces errors in depth judgement. In addition, Gruenefeld et al. (2020) found that techniques utilizing a grid help users estimate the distances to occluded objects more accurately. To achieve the goal of accurately locating occluded objects, the first design consideration of X-Board is that the assistive content needs to improve the perception of occluded objects. In reference to previous work, our current prototype of X-Board is designed as a translucent plane with grid markings, similar to a vertical chessboard, and located perpendicular to the ground between the eye of the user and the occluded object.

3.2.2 DP2. Visual coherence

Visual coherence refers to the coherence between real and virtual elements. The available definitions of AR visualization tend to focus not only on the mere fact that virtual data are mapped to visual representations but also on the spatial relationships and interactions between the physical world and virtual content (Kalkofen et al. 2011). However, AR content usually looks as if it is overlaid in front of any physical occlusions. To metaphorize the occlusion relationships, context data must also be a component of X-Board to support the combination of virtual data with the physical environment and the expression of the occlusion relationships between the X-Board and occluded objects. Avery et al. proposed an edge overlay visualization method that provides depth cues to make hidden objects appear to be behind walls (Avery et al. 2009). This method can be used to provide details about the occluding layers between the user and a remote location. Since our AR HWD platform (i.e. the HoloLens 2) has a spatial awareness function, spatial interactions can be considered as a source of context data. In our prototype, X-Board is implemented as a form of surface magnetism. A ray is cast from the AR HWD to the occluded object, and the X-Board is placed at the strike point on the occluding surface in the physical world. The X-Board switches its location as the user switches focus between target objects.

3.2.3 DP3. Adaptive display

An AR visualization method focuses on integrating virtual objects into the physical world. However, excess items, or inappropriate representations or organization thereof, can lead to a degradation in performance for some tasks, which is an important concern for emergency missions. Lu et al. (2020) proposed Glanceable AR, an information access paradigm for AR HWDs that includes two novel hands-free interfaces using head rotation or eye-tracked gaze data to access information. In addition to such explicit interaction, hidden assistance also plays an important role in AR. Lindlbauer et al. presented an optimization-based approach to automatically control when, where and how many virtual labels are shown (Sidenmark and Gellersen 2019). It is a real-time approach for automating this process based on the user's current cognitive load and knowledge about his or her task and environment. Alghofaili et al. investigated the use of gaze patterns for classifying the need for navigation aid in VR (Gruenefeld et al. 2019). Inspired by these hidden interaction methods, we adopted a data-driven method based on head and gaze data to display the X-Board which avoiding causing the user to miss important information in the physical world.

4 Implementation

4.1 X-Board visualization

Following the design principles identified above, we develop X-Board to help the user perceive multiple occluded objects better. To accurately illustrate occlusion and depth relationships, the X-Board is user-centric and suspended between the target and user, assigned a translucent mesh with grid markings. In this section, we propose two features of X-Board (i.e. surface magnetism and occluded target switching), which make it an acceptable and easy-to-use assistant. Notably, although it provides occlusion layer information and depth cues, it may also be obtrusive, obstructing the real environment and causing cognitive overload.

4.1.1 Surface magnetism

Context data are an important component, as they support the combination of virtual data with the physical environment (Zollmann et al. 2020). To give context to the content provided to the user, we define an invisible ray cast from the user's HWD to the target object and incorporate the Surface-Magnetism function of the Mixed Reality Toolkit (MRTK) in Unity to realize this feature. The X-Board is placed at and rotated around the strike point on the surface in the physical world in real time. As a result of this feature, the user can see through the X-Board to estimate the location of the occluded object when perceiving the situation in an indoor environment.

4.1.2 Occluded target switching

In our indoor perception tasks, multiple virtual avatars are displayed in the AR environment. However, the user can concentrate on only one avatar at a time, and the presence of too many X-Boards would obstruct the real environment and lead to heavy cognitive load. Therefore, for a concise implementation, only one X-Board is displayed in the surroundings of the user in the virtual environment, and the feature of occluded target switching is realized based on the user's head gaze direction. Our occluded target switching algorithm is summarized in Algorithm 1.

Algorithm 1 Occluded target switching based on head gaze

Input: Sets of occluded objects (O) and their locations (OL), the user's location (UL), the direction of the user's head gaze (D_{gaze}), and a direction threshold for switching (*Threshold*).

Output: The target object TO

1: **repeat for** each frame **do**
2: **for** $OL_i \in OL$ **do** $D_i \leftarrow OL_i - UL$
3: **for** $D_i \in D$ **do** $\Delta D_i \leftarrow D_{gaze} - D_i$
4: $\Delta D_{min} \leftarrow \mathrm{Min}(\Delta D)$
5: **if** $\Delta D_{min} <$ Threshold **then**
6: $j = \mathrm{Indexof}(\Delta D_{min})$
7: **return** $TO \leftarrow O_j$
8: else return *NULL*
9: end

In this algorithm, the sets of occluded objects (O) and their locations (OL) are generated for our indoor task. The user's location, UL, and the direction of the user's head gaze, $Dgaze$, are initialized and updated by the AR HWD. This algorithm runs every frame (line 1). First, each direction Di of a ray cast from the user to an object is computed (line 2), and the difference between the direction of each ray and the current direction of the user's head gaze is saved in a set ΔD (line 3). Then, the minimum difference $\Delta Dmin$ is found (line 4). If the minimum difference $\Delta Dmin$ is less than a specified threshold, then the target object TO will be replaced by the nearest object Oj (lines 5, 6, and 7); otherwise, the target object will not be changed (line 8).

4.2 Adaptive display based on gaze data

AR visualization is different from information visualization. The links between virtual and real elements, particularly spatial relationships and visual resource allocation, are given increased emphasis in AR visualization. In this section, we adopted an adaptive AR content display method based on real-time gaze data to balance the allocation of visual resources between virtual and real elements. Referring to previous work on adaptive navigation aids in VR (Gruenefeld et al. 2019), an LSTM network is employed to recognize the visual attention of the user (i.e. to classify whether the user wishes to observe the "nearby real environment" or "distant virtual objects") based on gaze sequence data.

4.2.1 Data collection and processing

To train the LSTM neural network for adaptive display, we have developed an AR toolkit to collect gaze sequence data with classification labels. The toolkit is implemented in the Unity engine. The participant can see both the virtual and real environments through the HoloLens 2. This state-of-the-art MR platform supports the function of user eye tracking. Figure 2 shows the first-person view with HoloLens 2. In Fig. 2, the hand gestures were projected on the wall by a projector, and they were control variables in our user evaluation that represented changes in real environments. Meanwhile, the virtual avatars displayed by HoloLens 2 represented the comrades which occluded by the wall.

Participants: We recruited 12 participants (4 female and 8 male) from a college to perform data collection. The participants ranged in age from 22 to 28 years old (μage = 24.50, STD = 1.42). Everyone reported normal or corrected-to-normal vision.

Procedures: First, each user was required to complete an eye-tracking calibration. This allowed the device to adjust the system for a more comfortable and higher-quality viewing experience for the user while ensuring

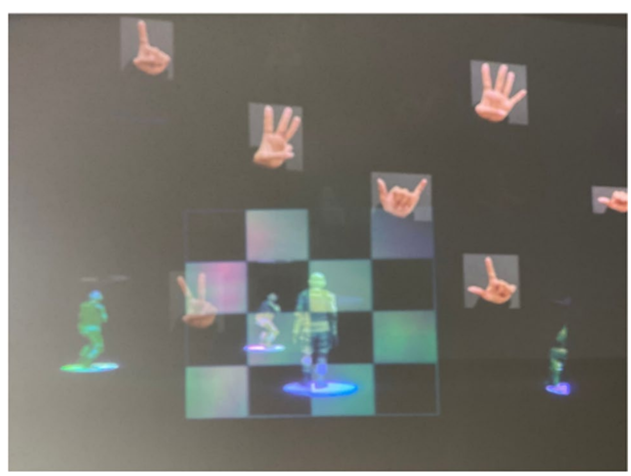


Fig. 2 First-person view of the toolkit for data collection

accurate eye tracking. After calibration, the participants were asked to focus their visual attention on either "the real environment" or "the virtual objects" and to adjust their visual attention as needed. The locations of the virtual objects were randomly generated at distances from the user ranging from 6 to 30 m in accordance with the indoor data collection environment. We logged each user's gaze rotation data and head gaze rotation data along with corresponding labels. All data were logged at the runtime frame rate (i.e. 60 Hz).

Data Collection: Benefitting from the sensors of the HoloLens 2, time series data on each user's eye gaze and head gaze could be effortlessly collected. To avoid any influence of the specific data collection environment, referring to previous work (Alghofaili et al. 2019), we also used the angle between the head gaze and eye gaze as input data. This angle is computed as follows:

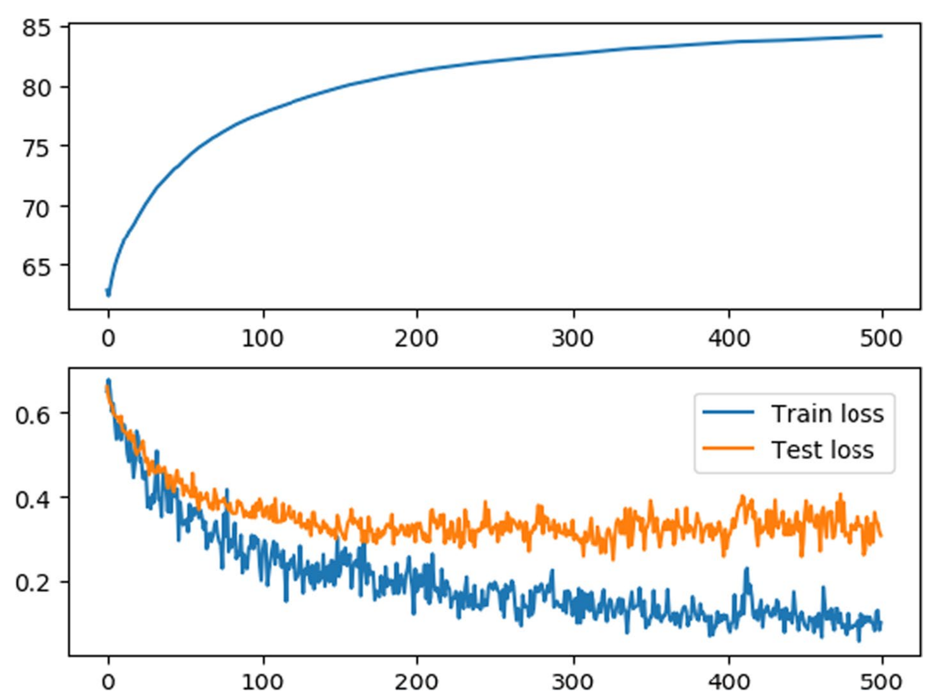
$$\theta = < \overrightarrow{Head_{gaze}}, \overrightarrow{Eye_{gaze}} >$$

where the θ is the angle between the head gaze and eye gaze vectors and is computed from the rotation data in Unity.

Data Processing: The remote mode of the HoloLens platform was used to collect the gaze data of the participants by means of a shortcut key; meanwhile, the corresponding labels ("nearby real environment" or "distant virtual objects") of the data were generated. To verify the effectiveness of the LSTM model, 20% of the fragments of a certain length from each set of data with each label were reserved as test data, and the rest were used for training. These fragments were selected randomly and isometrically as the input data. Figure 3 illustrates the time series data processing method.

Fig. 3 Gaze angle data processing

Fig. 4 Accuracy and loss values



4.2.2 Training and evaluation

We obtained 57,954 time series data with a window size of 150 frames (approximately 2.5 s) as input data.

Training: We employed a classical LSTM neural network with 1 hidden layer and 256 units for training on the binary classification task. The LSTM network was optimized using the Adam optimizer with the cross-entropy loss function. We trained our model for 500 epochs and obtained the best result in the 500th epoch.

Results: Fig. 4 shows the test accuracy achieved in each epoch and the loss values in the training and test epochs. The best result of this model was 84.16% accuracy on the test data, which is not a solid result. While there could be many error factors leading to this result, we believe that the most relevant factor is the labelling accuracy. It is difficult to ensure that participants will completely focus on a specific visual task. This is always a difficult problem when classifying the visual attention of humans, whether for a human or a machine. Therefore, as described in the next section, we adopted a mode filter to improve the accuracy for real-world application to perception tasks.

4.2.3 Mode filter for adaptive display

To improve the accuracy in real-world applications, we adopted a mode filter to process the time series results of the LSTM network. Accuracy and response speed are equally important. While the accuracy is basically proportional to the frame lengths, while the response speed is inversely proportional to the frame lengths. After many attempts and extensive analysis of the reference accuracy and the user experience of the response speed, a window size of 20 frames was chosen for the filter to be used in classification which gives consideration to both accuracy and response speed. With this filter, the label L at time t is computed as follows:

$$L_t = \begin{cases} Real, count(V)_{(t-20,t]} < count(R)_{(t-20,t]} \\ Virtual, count(V)_{(t-20,t]} \geq count(R)_{(t-20,t]} \end{cases}$$

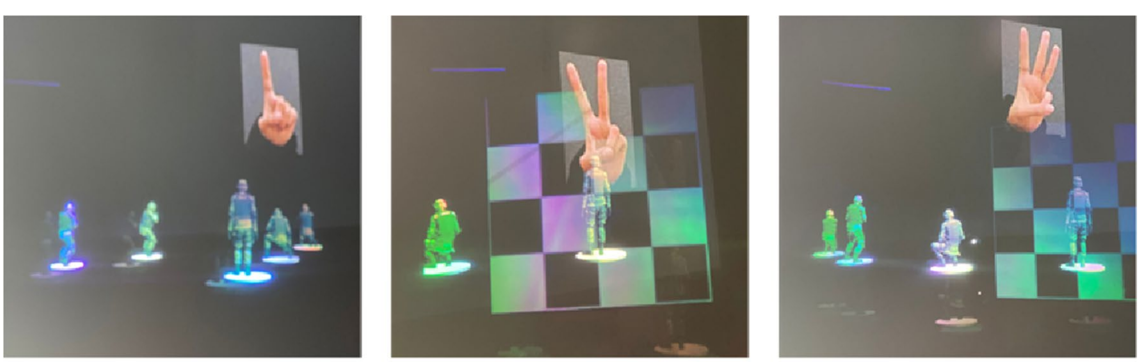
Table 1 shows the accuracy of the binary classification task with this filter was improved to 96.48% on the test data; moreover, less frequent switching was observed based on the users' feedback. Therefore, we adopted this method to realize the adaptive display feature of X-Board in order to balance the allocation of visual attention between virtual objects and the real environment.

Table 1 Accuracy of gaze data recognition

	LSTM	LSTM with mode filter
Accuracy	84.16%	96.48%

5 User evaluation

To explore the benefits and limitations of X-Board in user perception, we evaluated it in an empirical user study in a perception task. The aim of this user study was to compare the use of our egocentric adaptive AR assistant X-Board (Fig. 5c) with two other methods: without X-Board (Fig. 5a) and with X-Board but without the adaptive display feature (Fig. 5b). In Fig. 5, the hand gestures were projected on the wall by a projector, and they represented the changes in real environments. And the virtual avatars displayed by HoloLens 2 represented the comrades which occluded by the wall for the perception tasks. Accordingly, the participants were asked to perform a designed perception task using each of these three AR visualization methods. We investigated the effect of the X-Board on task completion time and perceptual accuracy. Additionally, we evaluated the usability of X-Board in regard to the task load of the participants by



a) without X-Board. b) with X-Board but without the adaptive display feature. c) with adaptive X-Board.

Fig. 5 First-person views of evaluation conditions asking them to complete a questionnaire. Additionally, user feedback was accounted for in the user study.

5.1 Participants

We recruited 16 participants (5 females and 11 male) from a college to conduct the test. The participants ranged in age from 22 to 28 years old (µage = 24.94, STD = 1.39). Five of them did not have prior experience with AR before the user evaluation, and most participants claimed to have good cognition of digital content ($\mu = 3.67$ on a 5-point Likert scale). Everyone reported normal or corrected-to-normal vision.

5.2 Apparatus

For the user study, we used the HoloLens 2 with a diagonal field of view of 52 degrees, a 2 k resolution and the application programming interface (API) for eye tracking enabled. We streamed AR content to the HoloLens instead of building and deploying a full project, allowing the app to take advantage of the more powerful resources of a PC. Moreover, the eye gaze data were communicated with the adaptive display module through the User Datagram Protocol (UDP), and the result of whether to display the X-Board was obtained by a module realized using the PyTorch library in real time. The perception task environment, which was simulated as an indoor anti-terrorism scenario, was developed in Unity 2019.4.28f with the MRTK toolkit provided by Microsoft.

To build the indoor environment in reality for the participants, we used a room with dimensions of approximately 4 m*9 m for this user study. To evaluate the user perception effect of X-Board in the perception task, in addition to the task accuracy and time, the movement distance of each participant was also recorded in consideration of the motion parallax. For this purpose, an optical tracking system (OptiTrack) was set up in the room. Figure 6a shows a diagram of the apparatus in the environment for the user evaluation. To overlay virtual objects on the real environment, we built a building-scale virtual environment centred on the user, and virtual avatars were placed randomly in the virtual environment. Figure 6b shows the virtual avatars in the Unity scene.

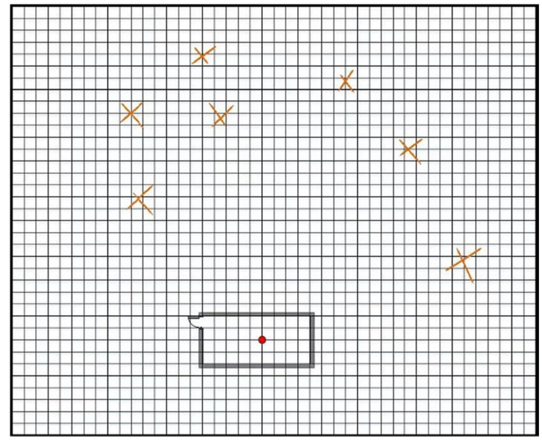


Fig. 7 Top-view map for marking the location of virtual avatars

The participants were asked to estimate the location of each virtual avatar and mark it on a map shown in Fig. 7. (depicting a top view of the room and a grid with distance markings) with a pad and pencil. We marked a coordinate origin in the room corresponding to the red point on the map. The room is centrally located on the map. The corresponding orientation and origin of the real room on the map would be told to the participants in advance. And the length of each square on the grid represents 1 m. The

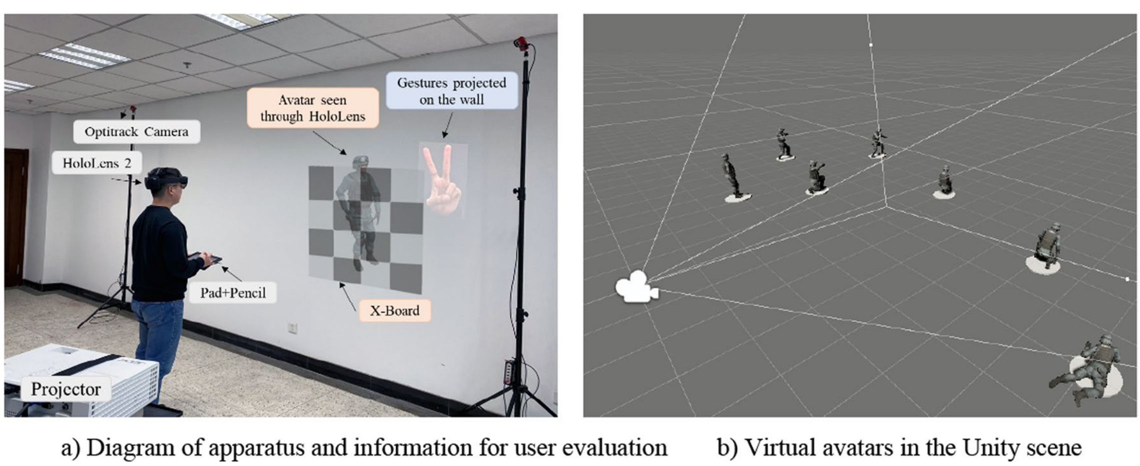


Fig. 6 Environment for the user evaluation

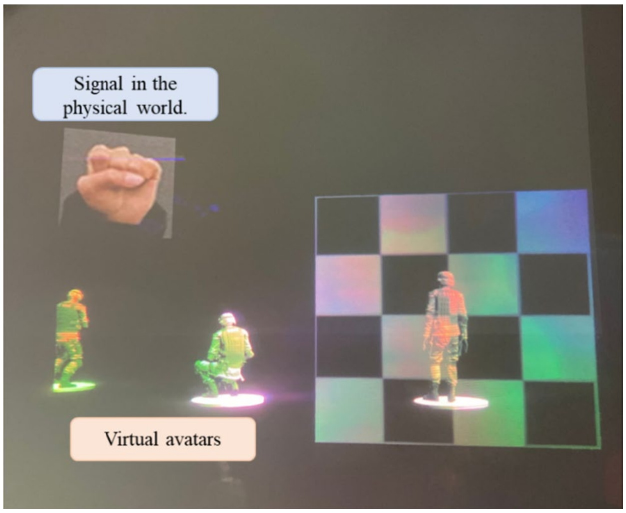


Fig. 8 Task for the user evaluation

participants' marked locations of estimates were recorded manually with a ratio conversion.

5.3 Tasks

A parallel perception task was adopted to evaluate the effect of X-Board. This task was composed of two parts: (a) watching and estimating the locations of virtual avatars and (b) watching and remembering signals given in the physical world. The first-person view is shown in Fig. 8.

5.3.1 Task a: watching and estimating virtual avatars' locations

In task a, the participants were asked to watch each virtual avatar representing an occluded comrade and to estimate their locations as quickly as possible. The virtual avatars were generated randomly in a certain distance range. In this task, 7 avatars were generated, and the distance from the user ranged from 6 to 30 m. A top-view map of the virtual building was provided to the participants. Participants were asked to mark the locations estimated on the map. During the user study, the participants were allowed to walk around in the room to estimate the locations of the virtual avatars based on the X-Board display and motion parallax, and their walking distances were tracked and recorded by the optical tracking system. In this task, the participants needed to focus their attention on distant virtual objects. However, there is a concern that a user who is wearing an AR HWD may ignore the nearby real environment, which is also important. Therefore, task b, which focused on the physical world, was required to be simultaneously executed in parallel.

5.3.2 Task b: watching and remembering signals given in the physical world

In task b, the participants were asked to watch and pay attention to the physical world and to remember dynamic signals. In this user study, the signals were represented as a series of gestures projected on the wall by a projector. We changed the type and location of the signal every 5–15 s. After the signal changed, there was a beep to remind the participants. After the participants had accomplished task a, they were asked to recall the signals given, and the accuracy achieved in this task was taken as an important factor in evaluating the usability of the adaptive display feature of X-Board. For this task, the participants needed to switch their attention from distant virtual avatars to nearby signals in the physical world. In such a set of parallel perception tasks, users with AR HWDs need to allocate their visual resources between virtual and real elements in an appropriately balanced manner to avoid missing key information in the physical environment. Such appropriate allocation of visual resources may be the most important feature of an AR assistant for use in real-world emergency tasks.

5.4 Measures

In this user evaluation, we focus on the user's perception in which AR content provides assistance. For objective evaluation with all three test conditions, the participants' performances (the locations of the estimated virtual avatars, the time of task completion, the distance of head movement, and the signals recalled by the participants) were logged during the user evaluation. The distance of head movement not only measures a participant's objective performance in the perception task but also enables the reconstruction of the task experience of all participants for a subjective evaluation. The accuracy of the locations of the estimated virtual avatars and the gesture signals are the key factors in evaluating the assistive perception effect of X-Board.

For subjective evaluation, we adopted the System Usability Scale (Bangor et al. 2008) (SUS) and the NASA Raw Task Load Index (Hart and Staveland 1988) (NASA RTLX) workload questionnaires to evaluate the usability and cognitive load with the X-Board. We also asked the participants to provide user feedback in this user study to fully discuss the usability of the X-Board.

We employed a within-subject design, so each participant performed tasks in all conditions. And the order of experiencing the three conditions was balanced across subjects using a Latin square design.

5.5 Procedures

Before the experiment, the participants were given 20 min to become familiar with the AR display style of the HoloLens. During the user study procedure, first, the participants were given a detailed introduction to the experimental background, the hardware, the three test conditions, and the perception task addressed in the study. Second, we helped the participants put on the AR HWD, and each participant was required to complete an eye-tracking calibration. For each test, the participants were asked to mark the locations of the virtual avatars on the map "as accurately as possible" and "as quickly as possible", with accuracy having a higher priority than speed. The participants were also informed that in the meantime, the signals given in the physical world should be noticed and remembered. Then, the participants completed each of the three test conditions one by one. When each condition was initialized, the locations of the virtual avatars were randomly generated with specified standard deviation and mean and normal distribution ($N(18, 12^2)$). Each time a participant completed the task, the participant's performances were logged. After finishing each designated experimental trial in the current condition, the participants were asked to fill out the SUS questionnaire and the NASA RTLX workload questionnaire to subjectively measure the cognitive load with X-Board. In the meantime, we asked them about their preferences and their feedback about X-Board. The study took approximately one hour for each participant.

5.6 Hypotheses

We tested three hypotheses in the user study ("**Baseline**": without X-Board; "**EG1**": assisted by X-Board without the adaptive display feature; "**EG2**": assisted by X-Board):

H1. For the part of the perception task that focuses on the virtual avatars' locations (task a), **EG1** will result in higher accuracy and greater efficiency than the **Baseline** condition since the user can perform estimation based on a reference object, and thus, the **EG1** condition will be preferred.

H2. For the part of the perception task that focuses on information in the physical world (task b), **EG1** will result in a lower accuracy of gesture signal than the **Baseline** condition due to the extensive occlusion caused by the X-Board display in the real world.

H3. For the parallel perception task (task a and task b), **EG2** will be preferred and will result in better performance, greater efficiency and less cognitive load than the two other conditions (**Baseline** and **EG1**) because of the grid mesh and the adaptive display feature, which will facilitate more balanced allocation of visual resources between the virtual and real elements and thus improve the perception ability.

6 Results

To test our hypothesis, we ran a series of data analyses. The results consist of two parts: an objective performance assessment and a subjective evaluation. Both of them will be presented in this section. The objective performance measure data were collected by user answers and sensors, and the subjective evaluation was conducted based on information gathered from questionnaires and user interviews. To understand how the results differed among the conditions ("**Baseline**": without X-Board; "**EG1**": assisted by X-Board without the adaptive display feature; "**EG2**": assisted by X-Board), we used repeated measures analysis of variance (RM-ANOVA) to interpret the objective measured data. A Greenhouse–Geisser correction was applied for violations of sphericity. For the qualitative data gathered from the questionnaires, we performed Friedman tests for the subjective measures. Bonferroni corrected pairwise comparisons are reported for significant main effects. We chose a significance level of 0.05.

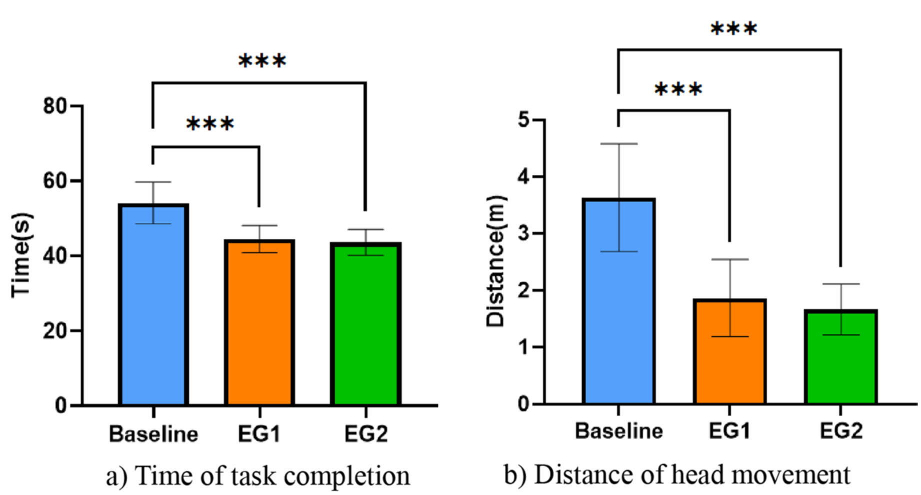


Fig. 9 Mean times and distances of users in the perception task

6.1 Objective performance

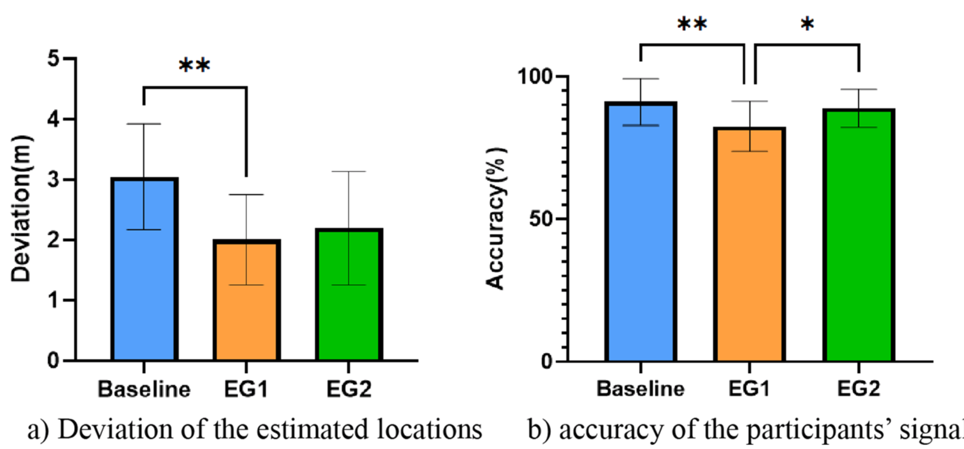
Figure 9a shows the times of task completion under the three conditions. For the perception task, **EG1** ($M=44.43$, $SD=3.52$) and **EG2** ($M=43.56$, $SD=3.35$) yielded lower times than the **Baseline** condition ($M=54.13$, $SD=5.38$). Our analysis of the main effect of task time showed statistically significant ($F_{1.875,\ 28.12}=57.73$, $p<0.001$). And pairwise analysis found that the differences were statistically significant between both the **Baseline** and **EG1** conditions ($p<0.001$) and the **Baseline** and **EG2** conditions ($p<0.001$). However, RM-ANOVA did not find a statistically significant difference in the task completion times between **EG1** and **EG2** ($p=0.99$).

Figure 9b shows the distances of head movement in the perception task under the three conditions. For the perception task, **EG1** ($M=1.86$, $SD=0.66$) and **EG2** ($M=1.67$, $SD=0.43$) resulted in lower head movement distances than the **Baseline** condition ($M=3.63$, $SD=0.92$). Our analysis of the main effect of the head movement distance showed statistically significant ($F_{1.926,\ 28.88}=52.63$, $p<0.001$). And pairwise analysis found that the differences were statistically significant between both the **Baseline** and **EG1** conditions ($p<0.001$) and the **Baseline** and **EG2** conditions ($p<0.001$). Similar to the results of task time, RM-ANOVA did not find a statistically significant difference in the distances of head movement between **EG1** and **EG2** ($p=0.94$).

Figure 10a shows the deviations of the estimated virtual avatar locations under the three conditions. For the perception task a, **EG1** ($M=2.00$, $SD=0.12$) and **EG2** ($M=2.19$, $SD=0.19$) resulted in lower deviations than the **Baseline** condition ($M=3.05$, $SD=0.20$). Our analysis of the main effect of the deviation showed statistically significant ($F_{1.460,\ 21.90}=9.099$, $p=0.002$). And pairwise analysis found that the differences were statistically significant between both the **Baseline** and **EG1** conditions ($p=0.0044$). However, RM-ANOVA did not find a statistically significant difference in the deviations of the estimated locations between the **Baseline-EG2** ($p=0.0516$) and **EG1-EG2** ($p=0.858$).

Figure 10b shows the accuracy of the participants' signal under the three conditions. For the perception task b, the **Baseline** ($M=91.03$, $SD=7.93$) and **EG2** ($M=88.89$, $SD=6.43$) conditions offered higher accuracy than **EG1** ($M=82.51$, $SD=8.48$). Our analysis of the main effect of the signal accuracy showed statistically significant ($F_{1.753,\ 26.30}=6.395$, $p=0.0071$). And pairwise analysis

Fig. 10 Mean results for the assistive effect in the perception task



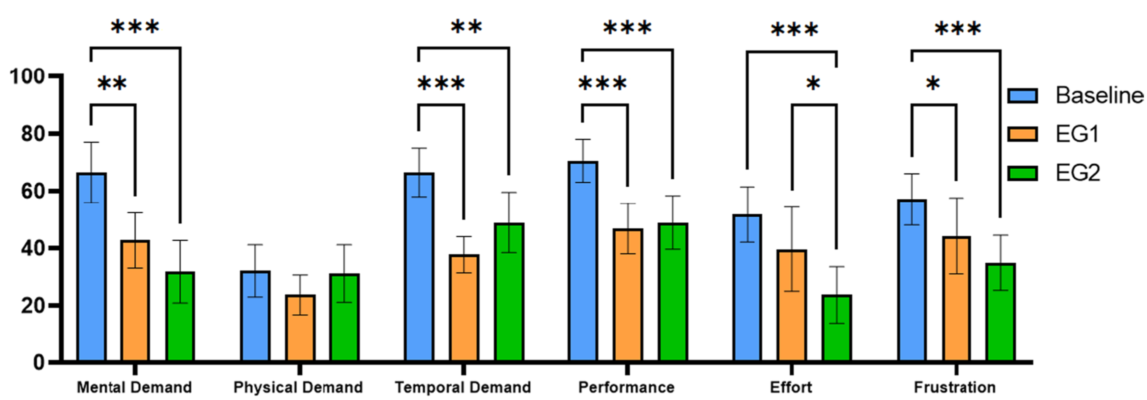


Fig. 11 NASA RTLX ratings

found that the differences were statistically significant between both the **Baseline** and **EG1** conditions ($p=0.0034$) and the **EG1** and **EG2** conditions ($p=0.049$). However, RM-ANOVA did not find a statistically significant difference in the participants' signal accuracy between the **Baseline** and **EG2** conditions ($p=0.99$).

6.2 Subjective evaluation

To evaluate the cognitive load caused by X-Board and the usability of the X-Boar, the NASA RTLX workload questionnaire and the SUS questionnaire were used to quantitatively evaluate the three test conditions. The user evaluation results in relation to our observations and the feedback received from the participants during the interview are also meaningful.

6.2.1 Results of the questionnaires

Figure 11 shows plot for the cognitive load in NASA RTLX questionnaire subscales Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration. All of the subscales showed similar effects as described in the following.

For Mental Demand, **EG1** ($M=42.81$, SD = 9.35) and **EG2** ($M=31.88$, SD = 10.59) obtained a lower rating than **Baseline** ($M=66.56$, SD = 10.27). There are statistical significances ($X^2=28.00$, $p<0.001$) between **Baseline-EG1** ($p=0.0023$) and **Baseline-EG2** ($p<0.001$). And we did not find statistical significance in the Mental Demand between **EG1** and**EG2** ($p=0.231$).

While for Physical Demand, **Baseline** ($M=32.19$, SD = 8.83), **EG1** ($M=23.75$, SD = 6.73) and **EG2** ($M=31.25$, SD = 9.76) obtained a similar rating. There were no statistical significances ($X^2=3.872$, $p=0.1443$) between each pair of these three conditions.

For Temporal Demand, **EG1** ($M=37.81$, SD = 14.19) and **EG2** ($M=49.06$, SD = 10.19) obtained a lower rating than **Baseline** ($M=66.56$, SD = 8.24). There are statistical significances ($X^2=29.42$, $p<0.001$) between **Baseline-EG1** ($p=0.0003$), **Baseline-EG2** ($p=0.0044$). However, there was no statistical significance in the Temporal Demand between **EG1** and**EG2** ($p=0.102$).

Furthermore, for Performance, **EG1** ($M=46.88$, SD = 8.45) and **EG2** ($M=49.06$, SD = 9.05) obtained a lower rating than **Baseline** ($M=70.63$, SD = 7.26). There are statistical significances ($X^2=21.43$, $p<0.001$) between **Baseline-EG1** ($p<0.001$), and **Baseline-EG2** ($p<0.001$).

Moreover, for Effort, **EG2** ($M=23.75$, SD = 9.60) obtained a lower rating than **Baseline** ($M=51.88$, SD = 9.33) and **EG1** ($M=39.69$, SD = 6.12). There are statistical

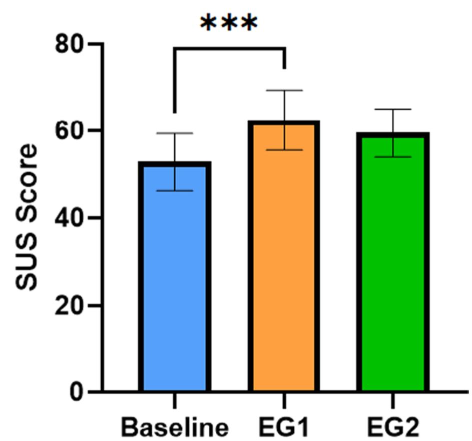


Fig. 12 SUS scores categorized by the test conditions

significances ($X^2=19.90$, $p= <0.001$) between **Baseline-EG2** ($p<0.001$), **EG1-EG2** ($p=0.0311$). However, we did not find statistical significance in Effort between **Baseline-EG1** ($p>0.2313$).

Lastly, for Frustration, **EG1** ($M=44.38$, SD = 12.85) and **EG2** ($M=35.00$, SD = 9.35) obtained a lower rating than **Baseline** ($M=57.19$, SD = 8.65). There are statistical significances ($X^2=21.29$, $p<0.001$) between **Baseline-EG1** ($p<0.001$), **Baseline-EG2** ($p=0.04$). And there was no statistical significances **EG1-EG2** ($p=0.1555$).

On the SUS questionnaire (Fig. 12), **EG1** ($M=62.5$, SD = 6.73) received a higher score than either the **Baseline** condition ($M=52.81$, SD = 6.37) or the **EG2** condition ($M=59.50$, SD = 5.39) in the perception task. There are statistical significances ($X^2=14.98$, $p<0.001$). **EG1** obtained a significantly higher score than the **Baseline** condition ($p<0.001$). No significant difference was found between the **Baseline–EG2** ($p=0.08414$) and **EG1-EG2** ($p=0.4719$).

6.2.2 User feedback

Here, we discuss the feedback received from the participants during the interviews, select some representative comments for analysis and divide these comments into three parts. Generally, the feedback collected from the participants shows that X-Board provides effective assistance for a perception task in an indoor environment.

AR-based environmental visualization method. All the participants thought that our AR-based perception assistance method was very useful for visualizing the locations of companions. ***P1:*** "Considering the difficulty of obtaining an indoor electronic map, I think it is an effective method to overlay and display the virtual avatars of companions directly using the AR method to improve users' perception ability". Some participants also did not deny the effectiveness of the AR method but felt that virtual avatars alone could not convey spatial information very well and that too much stacking of visual elements would impose too high a cognitive load. ***P3:*** "Virtual avatars are overlaid in my field of vision, but these graphics seem to have little connection with the real world, which makes it difficult for me to estimate the positions of these avatars in the real world".

X-Board visualization. The participants also reported that the X-Board could effectively assist in the perception task. ***P1*** and ***P6:*** "I think that the X-Board has a relative spatial relationship with the virtual avatar and the physical world, and it constructs a bridge between virtuality and reality, so that I can have a reference object to estimate the location of the virtual avatar." ***P4:*** "In the process of my experiment, I think the mesh of X-Board is helpful to estimate the distance, but it is not the main contributing factor". However, the participants also pointed out that the X-Board, which interacted with the physical world, obscured originally visible information. ***P2:*** "I find that I focus too much on the virtual graphics and ignore information in the real world, which may be dangerous in an emergency". ***P3, P7*** and ***P12:*** "I think the X-Board occludes the real information, makes it difficult for me to pay attention to the real world, and makes me feel the difficulty of the visual switch between virtual and real".

Adaptive display. We also received some positive comments on the adaptive display feature. ***P1:*** "I think I can effectively control whether the X-Board is displayed or not, which enables me to pay attention to nearby information while taking a glance at the virtual information." However, due to the unexplainability of deep learning, this feature sometimes suffered from display errors, which could confuse users. ***P7:*** "I think this feature is effective on the whole. Although some mistakes sometimes occur, I can adjust my visual attention well most of the time in the user evaluation experiment".

7 Discussion

We hypothesized that the **EG1** condition would result in higher accuracy and greater efficiency than the **Baseline** condition for the part of the perception task that focused on the virtual avatars' locations (**H1**). Our results supported **H1**. **EG1** yielded a lower deviation than the **Baseline** condition, and the difference between the **Baseline** and **EG1** conditions was statistically significant. We surmise that the most important factor in estimating the spatial locations of virtual avatars in physical space is "visual coherence". To perceive the current environment, users first need to use their visual perception ability to project virtual objects into the real world. This means that a connection needs to be established between virtuality and reality. As a virtual object, the X-Board intuitively fills this gap by interacting with the physical space. The NASA RTLX results show that X-Board significantly reduces the cognitive load compared with the **Baseline** condition. In addition, the user feedback shows that users tend to prefer the **EG1** condition over the **Baseline** condition. This indicates that the assistance of X-Board is effective for the task of estimating virtual avatar locations, and it is a technology that is easy for users to accept.

Our second hypothesis, **H2**, sought to confirm that the **EG1** condition will result in a lower accuracy of gesture signal than the **Baseline** condition due to the occlusion of real-world elements for the part of the perception task that focused on information in the physical world (**H2**). As expected, the gesture signal results indicate that the accuracy under the **Baseline** condition is higher than that under **EG1**. To some extent, this illustrates that the virtual graphics of AR may affect users' perception of the real world. Although AR can help emphasize some visual information, it may also introduce new problems because it obscures the real world.

We hypothesized that the **EG2** condition would be preferred over the other two conditions (**Baseline** and **EG1**) and result in better performance, greater efficiency and less cognitive load for the parallel perception task (**H3**). However, our results only partially supported **H3**. There was no significant difference between conditions **EG1** and **EG2** in terms of efficiency and accuracy in estimating the locations of the virtual avatars. Relative to **EG1**, **EG2** additionally includes the adaptive display feature. Therefore, we expected that condition **EG2** would enable higher accuracy in the visual perception of the real world. However, intelligent algorithms can introduce some problems. When the X-Board is displayed incorrectly, the user may be confused. However, most of the time, this function works normally. The results of the user evaluation show that this function can effectively assist in the perception task, improving the accuracy for the task and reducing the cognitive load on users. The NASA RTLX results show that there is no significant difference between the cognitive load of users in **EG2** and **EG1**, but the subscales are significantly lower than that in **Baseline**.

Overall, it appears that X-Board provided the most suitable assistance for the perception task. The **EG1** and **EG2** condition resulted in better performance, greater efficiency and less cognitive load than the **Baseline** condition. The participants' feedback concerning the adaptive display feature leads us to think about how artificial intelligence (AI) can be used to mine the user's intention to achieve more natural human–computer interaction. There still seems to be a contradiction between the unexplainability of AI and the deterministic feedback of human–computer interaction, which can cause confusion for users. Therefore, this data-driven assistance method for perception tasks needs to be further explored.

8 Limitations and future work

Our current AR-based assistance method suffers from several limitations. First, X-Board is a perception assistance technology based on the design principles proposed in this work. It is reasonable to believe that the more appropriate the display style is, the better the assistance effect. In our future work, the textures and materials of the X-Board will be related to the real-world environment to strengthen the link between virtuality and reality. Second, the real-time gaze-data-driven adaptive display method requires the collection of user data in the user evaluation environment, and there is still a chance of incorrect display results, which can confuse users. In the future, we intend to explore how real-time physiological data and AI might be combined to improve the efficiency of human–computer interaction. Third, according to the user feedback, there are some delays in the communication and rendering processes of the experimental equipment, resulting in drift of the virtual avatars, which affects some aspects of the usability of the prototype tool. Fourth, the randomly generated avatars for each test may cause random errors, due to the introduction of random variables. The experimental results always contain errors. Future we will design and adopt a more reasonable deployment of avatars (e.g. each condition corresponding to a location deployment), which avoids memory retention and minimize error.

Future, we will organize a more complex simulated SA task with a large and dynamic indoor scenario. With a complex SA task, we are going to measure the user's SA ability with all three levels (perception, comprehension, and projection), and use the specific metrics (e.g. SART or SAGAT (Endsley 2017, 2021) for direct SA measures.

9 Conclusion

In this research, we have proposed X-Board, an X-ray visualization-based AR assistant for perception in indoor environments. Based on our proposed design principles, X-Board provides visual–spatial cues by means of a grid mesh corresponding to the occluding surface in front of the target object. Meanwhile, the X-Board interacts with the physical world in real time, which improves the coherence between the virtual and real worlds. To ensure appropriate allocation of the user's visual resources, the user's visual intention is recognized based on gaze data to realize an adaptive display feature. We have evaluated X-Board in a perception task. The results of the user evaluation show that X-Board can effectively improve the accuracy and speed of perception and reduce the cognitive load on users and the adaptive display feature can avoid the case in which the user ignores information on the situation in reality while paying attention to the virtual information.

We believe that in the future, AR will become an important assistive technology for responding to emergencies. However, the act of overlaying a virtual display on a real scene, as one of the most important features of AR, will also introduce new problems in these task scenarios. For example, AR graphics may excessively attract users' visual attention, causing them to ignore important information in the real world. Design challenges still exist regarding how to enable users to achieve balanced allocation of visual resources in an AR environment. X-Board is an important step towards designing effective visual perception assistance mechanisms for future SA tasks.

Acknowledgements The authors wish to thank the anonymous reviewers for their constructive comments and the user evaluation participants for their time. This work was supported by the Pre-research Project of the 14th Five-Year Plan (Grant Number: 50904040201).

Data availability All the data and materials in the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest All authors of the paper make the following statement: No potential conflict of interest is reported by the authors. For research involving human participants, we collected data about their operations in every experimental environment and obtained their consent to use these data for academic purposes.

References

Alghofaili R, Sawahata Y, Huang H et al. (2019) Lost in style: gaze-driven adaptive aid for vr navigation. In: Proceedings of the 2019 CHI conference on human factors in computing systems, 1–12

Avery B, Sandor C, Thomas BH (2009) Improving spatial perception for augmented reality x-ray vision. In: 2009 IEEE virtual reality conference. IEEE, 79–82

Bane R, Hollerer T (2004) Interactive tools for virtual X-ray vision in mobile augmented reality. In: Third IEEE and ACM international symposium on mixed and augmented reality. IEEE, 231–239

Bangor A, Kortum PT, Miller JT (2008) An empirical evaluation of the system usability scale. Int J Hum–comput Interact 24(6):574–594

Barnum P, Sheikh Y, Datta A et al. (2009) Dynamic seethroughs: synthesizing hidden views of moving objects. In: 2009 8th IEEE international symposium on mixed and augmented reality. IEEE, 111–114

Dey A, Cunningham A, Sandor C (2010) Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments. In: Proceedings of the 17th ACM symposium on virtual reality software and technology, 211–218

Dey A, Jarvis G, Sandor C et al. (2011) An evaluation of augmented reality x-ray vision for outdoor navigation. Virtual Real Society of Japan

Diaz C, Walker M, Szafir DA et al. (2017) Designing for depth perceptions in augmented reality. In: 2017 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, 111–122

Dominic J, Robb A (2020) Exploring effects of screen-fixed and world-fixed annotation on navigation in virtual reality. In: 2020 IEEE conference on virtual reality and 3d user interfaces (VR). IEEE, 607–615

Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. Hum Factors 37(1):32–64

Endsley MR (2017) Direct measurement of situation awareness: Validity and use of SAGAT. Situational awareness. Routledge, London, pp 129–156

Endsley MR (2021) A systematic review and meta-analysis of direct objective measures of situation awareness: a comparison of SAGAT and SPAM. Hum Factors 63(1):124–150

Erat O, Isop WA, Kalkofen D et al (2018) Drone-augmented human vision: exocentric control for drones exploring hidden areas. IEEE Trans Visual Comput Graph 24(4):1437–1446

Eren M T, Cansoy M, Balcisoy S (2013) Multi-view augmented reality for underground exploration. In: 2013 IEEE virtual reality (VR). IEEE, 117–118

Furmanski C, Azuma R, Daily M (2002) Augmented-reality visualizations guided by cognition: perceptual heuristics for combining visible and obscured information. In: Proceedings. International symposium on mixed and augmented reality. IEEE, 215–320

Gagnon HC, Rosales CS, Mileris R et al (2021) Estimating distances in action space in augmented reality. ACM Trans Appl Percept (TAP) 18(2):1–16

Gebhardt C, Hecox B, van Opheusden B et al. (2019) Learning cooperative personalized policies from gaze data. In: Proceedings of the 32nd annual ACM symposium on user interface software and technology, 197–208

Gruenefeld U, Prädel L, Heuten W (2019) Locating nearby physical objects in augmented reality. In: Proceedings of the 18th international conference on mobile and ubiquitous multimedia, 1–10

Gruenefeld U, Brück Y, Boll S (2020) Behind the scenes: comparing X-ray visualization techniques in head-mounted optical see-through augmented reality. In: 19th international conference on mobile and ubiquitous multimedia, 179–185

Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol North-Holl 52:139–183

Hertel J, Steinicke F (2021) Augmented reality for maritime navigation assistance-egocentric depth perception in large distance outdoor environments. In: 2021 IEEE virtual reality and 3D user interfaces (VR). IEEE, 122–130

Kalkofen D, Sandor C, White S et al (2011) Visualization techniques for augmented reality. Handbook of augmented reality. Springer, New York, pp 65–98

Kytö M, Mäkinen A, Häkkinen J et al (2013) Improving relative depth judgments in augmented reality with auxiliary augmentations. ACM Trans Appl Percept (TAP) 10(1):1–21

Lilija K, Pohl H, Boring S et al. (2019) Augmented reality views for occluded interaction. In: Proceedings of the 2019 CHI conference on human factors in computing systems, 1–12

Lindlbauer D, Feit AM, 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, 147–160

Liu JM, Narasimham G, Stefanucci JK et al. (2020) Distance perception in modern mobile augmented reality. In: 2020 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW). IEEE, 196–200

Livingston MA, Ai Z, Karsch K et al (2011) User interface design for military AR applications. Virtual Real 15(2–3):175–184

Livingston MA, Swan JE, Gabbard JL et al. (2003) Resolving multiple occluded layers in augmented reality. In: The second IEEE and ACM international symposium on mixed and augmented reality, 2003. Proceedings. IEEE, 56–65

Livingston MA, Ai Z, Swan JE et al. (2009) Indoor vs. outdoor depth perception for mobile augmented reality. In: 2009 IEEE virtual reality conference. IEEE, 55–62

Lu F, Davari S, Lisle L et al. (2020) Glanceable ar: evaluating information access methods for head-worn augmented reality. In: 2020 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, 930–939

Minaskan N, Pagani A, Dormoy CA et al. (2021) A study of human-machine teaming for single pilot operation with augmented reality. In: 2021 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-adjunct). IEEE Computer Society, 397–402

Nilsson S, Johansson B, Jonsson A (2009) Using AR to support cross-organisational collaboration in dynamic tasks. In: 2009 8th IEEE international symposium on mixed and augmented reality. IEEE, 3–12

Osmers N, Prilla M. Getting out of out of sight: evaluation of AR mechanisms for awareness and orientation support in occluded multi-room settings. In: Proceedings of the 2020 CHI conference on human factors in computing systems, 1–11

Pascale MT, Sanderson P, Liu D et al (2019) The impact of head-worn displays on strategic alarm management and situation awareness. Hum Factors 61(4):537–563

Peillard E, Argelaguet F, Normand JM et al. (2019) Studying exocentric distance perception in optical see-through augmented reality. In: 2019 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, 115–122

Pfeuffer K, Abdrabou Y, Esteves A et al (2021) ARtention: a design space for gaze-adaptive user interfaces in augmented reality. Comput Graph 95:1–12

Ping J, Liu Y, Weng D. Comparison in depth perception between virtual reality and augmented reality systems. In: 2019 IEEE conference on virtual reality and 3D user interfaces (VR). IEEE, 1124–1125

Sandor C, Cunningham A, Dey A, et al. (2010) An augmented reality X-ray system based on visual saliency. In: 2010 IEEE international symposium on mixed and augmented reality. IEEE, 27–36

Sharma A, Nazir S, Ernstsen J (2019) Situation awareness information requirements for maritime navigation: a goal directed task analysis. Saf Sci 120:745–752

Sidenmark L, Gellersen H (2019) Eye&head: synergetic eye and head movement for gaze pointing and selection. In: Proceedings of the 32nd annual ACM symposium on user interface software and technology, 1161–1174

Swan JE, Singh G, Ellis SR (2015) Matching and reaching depth judgments with real and augmented reality targets. IEEE Trans Visual Comput Graph 21(11):1289–1298

Tsuda T, Yamamoto H, Kameda Y et al (2006) Visualization methods for outdoor see-through vision. IEICE Trans Inf Syst 89(6):1781–1789

Uratani K, Machida T, Kiyokawa K et al. (2005) A study of depth visualization techniques for virtual annotations in augmented reality. In: IEEE Proceedings. VR 2005. Virtual reality. IEEE, 295–296

Vaziri K, Bondy M, Bui A, et al. (2021) Egocentric distance judgments in full-cue video-see-through VR conditions are No better than distance judgments to targets in a void. In: 2021 IEEE virtual reality and 3D user interfaces (VR). IEEE, 1–9

Winter S, Tomko M, Vasardani M et al (2019) Infrastructure-independent indoor localization and navigation. ACM Comput Surv (CSUR) 52(3):1–24

Wither J, Hollerer T (2005) Pictorial depth cues for outdoor augmented reality. In: Ninth IEEE international symposium on wearable computers (ISWC'05). IEEE, 92–99

Zhang J, Xia X, Liu R et al (2021) Enhancing human indoor cognitive map development and wayfinding performance with immersive augmented reality-based navigation systems. Adv Eng Inform 50:101432

Zollmann S, Langlotz T, Grasset R et al (2020) Visualization techniques in augmented reality: a taxonomy, methods and patterns. IEEE Trans Visual Comput Graph 27(9):3808–3825

Zollmann S, Kalkofen D, Mendez E et al. (2010) Image-based ghostings for single layer occlusions in augmented reality. In: 2010 IEEE international symposium on mixed and augmented reality. IEEE, 19–26

Zollmann S, Grasset R, Reitmayr G et al. (2014) Image-based X-ray visualization techniques for spatial understanding in outdoor augmented reality. In: Proceedings of the 26th Australian computer–human interaction conference on designing futures: the future of design, 194–203