Research Article

Classification, Application, Challenge, and Future of Midair Gestures in Augmented Reality

Yi Lu,1 Xiaoye Wang,1 Jiangtao Gong,2 Lejia Zhou,1 and Sen Ge1

1Beijing University of Technology, Beijing, China
2Institute for AI Industry Research (AIR), Tsinghua University, Beijing, China

Correspondence should be addressed to Yi Lu; luyi@bjut.edu.cn

Received 31 August 2022; Accepted 22 September 2022; Published 5 October 2022

1. Introduction

Augmented reality (AR) allows users to perform tasks in both real and virtual environments by overlapping virtual content in the real world [1]. AR is featured by real-time interaction, where gesture, voice, body posture, and even eye movements can be interaction mode. Gesture with wide application in human-computer interaction (HCI) can be used as input media for various media, including mobile phones, computers, televisions, large screen displays, etc. Its interaction can be direct touch, using physical devices (e.g., pens, remote controls, and handles) and body posture or their combinations. Gestures not using any physical medium are called midair gestures. Compared with using handle and touch, midair gesture realizes nonphysical contact with computer and manipulates objects in virtual environment with less constraints. In AR, natural midair gestures provide an intuitive interactive method to connect the virtual and real worlds. More natural and flexible than handle and buttons, midair gesture interaction can intuitively demonstrate information and express intention, which reduces the user’s learning cost, and greatly enhances immersive interaction [2, 3].

There are a variety of midair gestures, and the gesture type may vary with the application environments. According to different time, use environment and functions, the classification methods of gestures are also different. This paper will explain the different classifications from the perspective of user experience. The change of gesture classification is also closely related to the development of gesture recognition technology. Gesture recognition is the key to midair gesture interaction. It is a technology to operate the device by capturing human’s limb movements and converting them into corresponding commands [4]. Currently, gesture based systems such as Microsoft Kinect and Leap Motion, which support hand and whole body tracking, have become ubiquitous. Midair gesture has become a popular interaction mode with the emergence of more and more systems such as Microsoft Hololens and Magic Leap [5]. In the design of midair gestures, the core is to provide a comfortable user experience. Gesture type and gesture recognition technology are important factors that affect the user experience, and also determine the application of midair gesture. Therefore, midair gesture interaction in AR has important research value. The future user experience design of midair gesture will be benefited from the understanding of its current design,
application, and technical development. To this end, the recent development of midair gesture in AR is combed and analyzed, and the future development trend of midair gesture is discussed, with the focus on the following aspects: (1) different classification methods of midair gestures; (2) typical application fields of midair gestures; (3) challenges and future development of midair gesture interaction.

2. Classification and Elicitation Study of Midair Gestures

2.1. Classification of Midair Gestures

2.1.1. Classification in HCI. Gesture is often used for assisting verbal communication, and the research on midair gesture also originates from the observation of daily interpersonal communication. Therefore, the earliest classification of midair gestures was also based on speech [6–8]. The midair gesture based on the language environment provides a basis for the classification of gestures in HCI in the future. The interface combining gesture and voice is explored through human to human communication. Researchers have also done a lot of work on the interface of simple midair gesture control. It has always been a hot issue in HCI. Karam and Schraefel created a wide range of taxonomies and proposed a model of deictic, schematic, manipulation, signal, and sign language gestures. In this model, gestures are also equivalent to symbolic or descriptive gestures, which are used to describe body shape and the form indicated by speech [9]; Aigner et al. extended their model, and found that pointing, pantomime, direct manipulation, signal gesture, and icon gesture are the most popular gestures by analyzing 5500 gestures [10]. Moreover, gestures were further divided into static gestures and dynamic gestures according to their temporal state, and signal gestures were further divided into Semaphoric Static, Semaphoric Dynamic, and Semaphoric Stroke. Icon gestures were divided into Iconic Static and Iconic Dynamic. Point gesture is used to indicate the object and direction. Semaphoric- Static and Semaphoric-Dynamic are related to the social culture we are familiar with. For example, in daily life, we thumb up to mean “OK”. Unlike Semaphoric-Static and Semaphoric-Dynamic, Semaphoric-Stroke has special meanings in some cases, especially in controlling multimedia, such as palm waving to switch to the next page. Pantomonic gesture is a more vivid performance action or continuous action. Icon-Static and Icon-Dynamic are related to images and icons, performing specific tasks with a shape (circle, triangle, etc.) made by hand. Manipulation gesture is also an abstract gesture used to manipulate objects, such as zoom in and out.

2.1.2. Classification Based on Time. According to the dimension of time, midair gestures can be classified into static gestures and dynamic gestures. Static gesture is the specific state of the hand including the posture, position, and direction of the hand. Dynamic gesture is a time series of hand states, which means that the posture, position, and direction of dynamic gesture will change with time. Because dynamic gesture has a higher degree of freedom and more operable actions, it is more difficult to recognize in gesture recognition. Nevertheless, dynamic gesture can be recognized more easily with the maturity of gesture recognition technology.

2.1.3. Classification Based on Definition Mode. Midair gestures are classified into predefined and user-defined gestures from the perspective of technology and experience. Predefined gestures mean that system engineers have designed gestures according to one’s usage habits for the use of systems or some applications before we use some system devices, and users have to learn to use these gestures. User-defined gestures are usually gesture types defined by users according to their own preferences, which requires more complex gesture recognition technology. For example, Wang et al. have created a gesture AR authoring system, in which users complete freehand interaction using a visual programming interface to match gestures with the response of AR virtual content. It improves the flexibility and independence of freehand interaction [11]. However, more researchers are inclined to study user-defined gestures with the continuous development and maturity of gesture recognition technology. This method has many advantages. Users do not have to learn before using the system, and the interaction is more natural. Moreover, it is also consistent with the user-centered design principle in HCI.

2.1.4. Classification Based on Task Type. The classification of gestures is closely related to their purpose and use environment. Communicative gestures are often suitable for performing some “abstract” tasks, such as controlling menus, turning on/off devices, media control tasks, especially signal gestures. According to Groenewald et al., midair gestures are mainly used for selection, navigation, and operation tasks. Manipulation gesture is also suitable for “physics based interaction” operation tasks in addition to navigation tasks. It can be interpreted as operating virtual objects like real objects, providing users an immersive experience that they are directly operating something, like playing a virtual basketball with tapping gestures, throwing a paper plane with kneading actions, etc. Generally, a more immersive user experience can be created with the combination of “physics-based interaction” and sensory effects such as vision and hearing in the AR environment with the combination of virtuality and reality. Therefore, midair gestures can be divided into abstract gestures and gestures based on physical operations according to their current application in AR. Abstract gestures are used to perform operations like menu selection, mode switching, switching, and object adjustment. Such gestures are also related to culture and habits. Of course, they also convey some special semantics. For example, swinging the palm indicates the state of fish swimming. Gestures based on physical operations are closely related to daily life, including grasping, kneading, tapping, etc.

At present, the exploration of gesture applications is also constantly innovated based on a variety of gesture recognition technologies, shifting from leisure AR applications to deeper areas. Abstract gestures are often ambiguous, and it is a challenge for users to learn specific actions. To this end, some studies have focused on how to use simple gestures to operate a variety of tasks, thus reducing users’ learning burden while enhancing the user experience. Designers
from Leap Motion are exploring a one-hand operation scheme suitable for AR, using simple kneading gestures to execute a variety of quick commands. Besides, they designed three gesture modes, i.e. throwing mode, slingshot mode, and time mode to test the kneading gesture scheme. They hope to create more one-hand control modes, and enable each hand to perform a variety of tasks by combining with two-hand gestures.

2.2. Elicitation Study. As gesture recognition technology has become mature with diversified applications, relevant gestures are also in large number. Therefore, later studies have also focused more on evaluating different sets of midair gestures to know the effective gestures for specific operation tasks.

Researchers need to evaluate the most comfortable gestures for users, so as to explore the gestures with user preference in AR. The most common method is gesture elicitation study. Gesture elicitation is a technology widely used to recognize self-discoverable gesture words in HCI [12]. Usually, the participants are shown the reference (effect), and then they are asked to propose a more matching (easy and intuitive) gesture. Researchers classify a large number of data after collecting a large number of gestures, so as to obtain the gestures with user preference. However, such research is limited to hand and finger movements using specific technologies (e.g., public displays, TV, AR, and VR) [12], such as using midair gestures to control TV media. Figure 1 shows the set of midair gestures determined by Samimi et al. for TV presenters through elicitation study. The gesture set consists of five gestures from two camera shots (long shot and close shot). The results of the evaluation study show that the derived set of gestures will not consume too much physical strength and attention of the host. The host using these gestures are enabled to control the AR content in the TV and tell stories in a modern way with more power of expression. Figure 2 shows the usage scenario [13].

Wobbrock et al. developed a set of user-defined gestures according to the degree of consensus among participants, and classified the induced gesture vocabulary. This classification method aims to expand the gesture design space in the desktop environment [14]. Lee conducted a Wizard of Oz study on the AR multimodal interface, aiming to explore the types of gestures people want to use in the tasks of virtual object manipulation. The experimental results showed that the most popular gesture types are pointing, translation, and rotation gestures [15]. Piumsomboon et al. explored the types of natural gestures in AR through experiments and determined the gesture sets corresponding to the four basic tasks of select all, open, close, and select horizontal menu. They got 800 user-defined gestures in 40 tasks, and finally got 44 user-defined gesture “consensus sets”, providing an important reference for designers of user experience. This gesture set represents the gesture preferences of users in AR tasks [16]. Moran-Ledesma et al. utilized heuristic research methods to determine the gesture set in which physical props are used to control VR and carried out necessary construction on the basis of previous work, so as to include the selection of both gestures and physical objects used together [12].

3. Interactive Application of Midair Gesture for AR

3.1. Midair Gesture in Multimedia. Many researches focus on gesture control of various types of media, such as video, virtual content playback, and so on. The interaction with

| Gesture Mode | Forward | Backward | Start | Stop |
|--------------|---------|----------|-------|-------|
| Close shot   | ![Close shot gesture](image1) | ![Close shot gesture](image2) | ![Close shot gesture](image3) | ![Close shot gesture](image4) |
| Long shot    | ![Long shot gesture](image5) | ![Long shot gesture](image6) | ![Long shot gesture](image7) | ![Long shot gesture](image8) |

**Figure 1**: The gesture set determined by Samimi et al. according to the evaluation results.

**Figure 2**: Usage scenario of midair gesture (Samimi et al.).
users in applications is not limited to mobile electronic devices. In the AR vehicle interaction system, the displayed content can be switched through midair gestures. Outdoor AR advertisements trigger the playback of virtual content through the waving and staying of gestures. Rovelo Ruiz et al. [17] introduced a set of gestures for controlling panoramic videos. In addition, Siddhpuria et al. [18] explored the use of discrete micro gestures with smart watches to control remote media. And several studies used midair gesture interaction to improve the user experience of smart TV, especially multimode technology and interaction [19, 20]. Rateau et al. [21] proposed user-defined gestures to define virtual interaction space in a universal environment.

3.2. Embodied AR Presentation. Midair gesture interaction has shown many advantages in experimental dynamic simulation, three-dimensional object display, and other aspects. Most systems predefine the virtual content triggered by each gesture—create mapping relationships. Besides, users need to learn these gestures in advance, and the execution of predefined gestures will trigger predefined actions. For example, Chalktalk VR/AR is a simulation tool for creating drawings in face-to-face brainstorming. Users can call vivid and intelligent virtual elements through aerial drawing to explain scientific knowledge online. All triggers need to be preprogrammed [22]. In the Post-Post-its system, researchers designed a series of natural and near-realistic midair gestures to move, copy, delete, and virtual post, and it is used for brainstorming in online classes [23]. Wang et al. created the GesturAR on Hololens2 using Unity3D. Users can complete freehand interaction by using the visual programming interface to match gestures with the response of AR virtual content [11]. Unlike other works, users design gestures to interact with virtual elements completely on their own in GesturAR, which improves the flexibility and independence of freehand interaction (Figure 3).

3.3. Midair Gesture in Teaching Presentation. Midair gesture interaction based on AR demonstration can intuitively display information, so it has great potential in performance, speech, and teaching. Vision based real-time gesture recognition has the advantages of low learning cost, noncontact control, richer, and more natural interactive actions. Therefore, gestures often replace the keyboard and mouse to complete the basic virtual interaction function in the teaching scene [2]. For example, predefined gestures are mapped to virtual interactive commands, and the multimedia platform is operated with midair gestures such as confirm, return, select, grab, and release. Users do not need other auxiliary tools by combining the demonstration, entertainment, and teaching of multimedia technology with gesture recognition technology, and a natural direct and humanized HCI experience is achieved with the use of gestures.

Saquib et al. proposed a dynamic demonstration tool that directly operates the virtual interface using the Leap Motion. Users can call virtual graphic elements in real time through body posture, so as to use daily actions to enhance the ability of communication with the audience [24] (Figure 4).

At present, many AR education use HMDs and holographic projection to create teaching environments, but Gong et al. created Holoboard base on pseudo holographics. They designed and implemented a rich set of novel interaction technologies, including body posture interaction, gesture and handle interaction, and tactile feedback, so that teachers and users can achieve naked eye augmented reality teaching experience through immersive demonstration, role-playing, and behind the scenes lectures [25].

Another typical application is the midair gesture operation AR experiment. With the help of multimedia, simulation, AR, and other technologies, the relevant software and hardware operating environment that can assist the operation links of traditional experiments is created on the computer, and the experimenter can complete various experiments as in the real environment. An example is the immersive chemical experiment system with head-mounted displays and Leap Motion gesture input devices. Learners are free to grab, drag, and drop experimental instruments according to the actual operation mode to complete the experimental operation. In addition, Geping et al. discussed the impact of experiments based on gesture interaction technology on learners’ experience, and found that virtual experiments based on gesture interaction technology can effectively enhance learners’ immersive experience, thereby increasing their learning motivation [26].

The application of midair gesture in teaching is not limited to virtual experiments, but can also realize the 6DOF tracking effect with good stability and high accuracy with the help of AR glasses and monocular and multicamera. It can complete the operations of two hand linkage, midair triggering, dragging, and moving, and realize the natural and flexible interactive experience of midair gesture.
In the field of education, midair gesture interaction has changed the traditional teaching method, made the teaching content expressed in a more wonderful way, and enhanced the interactivity and entertainment in the teaching process. Different from the application in other fields, teaching demonstration is not only personal experience but also expression and communication. Therefore, when using head mounted display, the personal experience effect is good, but the communication effect may be poor. So technically, using Leap Motion and Kinect for tracking and recognition will be more suitable for use in teaching situations, because it can help teachers avoid additional physical burden and will not restrict communication. The application of midair gesture in the field of education and teaching is of great value. The reason is that the AR teaching environment itself has several particularities:

1. **Environmental Diversity.** The environment configuration of online and offline classes is different. In offline classes, teachers move near the podium, but because their body posture often shifts between facing the blackboard and facing the students, and the radial rotation is omnidirectional. In this process, it is easy to cause false recognition due to occlusion. It is necessary to consider how to use gesture recognition equipment to accurately track midair gestures. In remote classes, the interaction range is limited, and it is also required to consider the way of gesture recognition and how to configure the camera so that gestures can display a full range of presentation contents in a limited range. In addition, the interaction between teachers and students in the online classes has also become difficult, which will also affect the user's interactive experience.

2. **Multiplicity of Interactive Content.** Blackboard writing and speech contain rich contextual information. Different from pure single user interaction, teachers are constantly interacting with students through teaching content while interacting with AR space. It is necessary to consider equally individual user experience and demonstrating and imparting knowledge to others. This makes the task of gesture interaction more complex. How to improve user experience and teaching efficiency will be the key research content of the application of midair gesture interaction in teaching presentation.

3. **High-Frequency Use of Gestures and Large Number of Meaningless Actions.** Most teachers have accumulated a considerable number of unconscious gestures in their long-term teaching work. These gestures vary from person to person and often only express the feelings of teachers when expressing knowledge, with little relevance with the knowledge itself. These gestures are likely to be recognized by the system as primitives with the ability to activate interactive tasks. Such feature of users like teacher imposes higher requirements for interaction design. Because the midair gesture belongs to the user interface that is always in the "on" state, the system needs to distinguish between user actions that drive interaction tasks and those that are only unconscious operations. It is necessary to prevent the situation in which users can do nothing about the equipment by making full use of the situational information in the teaching state.

Midair gesture interaction has broad prospects in teaching. Especially because the epidemic control and prevention become normalized, more students have to take online classes. Compared with the traditional boring video online classes, AR allows teachers to interact with students while demonstrating through midair gestures. In addition, it realizes the demonstration of the surgical process by multiperson remote cooperation in the surgical training teaching. In some skills training, engineers can also demonstrate mechanical assembly through midair gestures. Through the
remote auxiliary function, learners get remote virtual guidance.

4. Challenges of Midair Gesture Interaction

4.1. Gesture Recognition. Midair gesture interaction depends on gesture recognition. For midair gestures, the mainstream recognition method is gesture recognition based on computer vision. Gesture recognition is a perceptual computing user interface, which enables the computer to capture and interpret gestures and execute commands according to the understanding of a gesture [27]. Gesture recognition usually includes the following steps: first, acquire the gesture frame, then, track the gesture, extract the features (finger state, thumb state, skin color, alignment, finger, and palm position) [27], and finally, classify to get the output gesture. As shown in Figure 5.

Gesture acquisition is to capture human gesture images by computer [28], which can be achieved with vision based recognition. Webcams or depth cameras can be used in the absence of needs for special equipment. In addition, special tools can also be used, such as motion sensing and input devices that capture hand gestures and motion (Microsoft Kinect, Leap motion, etc.) [29]. Researchers use a portable device, Leap Motion to achieve a full hand skeleton and perform object operations with higher accuracy [30]. In the study of gesture interaction only, researchers have explored a variety of methods to carry out the best gesture recognition for specific gestures. Lee et al. [31] designed gloves with conductive fabric on fingertips and palms for gesture recognition, using vibrating motors for tactile feedback and marks around the wrist for tracking fingers and also designed some gestures for selection, grasping, cutting, and copying. Lee and Hollerer created HandyAR, a system enabling freehand interaction with standard webcams. Its gestures supported are limited to the open/closed hand for object selection and hand rotation for object inspection [32]. Their subsequent work allowed the relocation of objects using unmarked tracking [33]. Fernandes and Fernandez used hand images to train statistical models to allow freehand detection [34]. And FingARtips [35] enabled users to knead and move virtual contents by detecting benchmarks on their fingertips. The 3D geometry of hand is retrieved for collision detection of hand virtual objects with the development of RGB-D and stereo camera [36, 37]. Although the air gesture interaction based on computer vision, which does not require additional equipment, is more natural, the accuracy of recognition is still limited by many factors. So most previous work mapped the specified gestures to limited operations, namely selection, translation, or rotation, without considering their easy and natural operation. In other words, gesture detection is only used as a substitute function of the mouse to operate 3D content, and the flexibility of the hand has not been brought into full play. Therefore, when performing gesture interaction, we need to consider the needs of specific interaction scenarios and tasks, so as to improve the user interaction experience.

Gesture recognition is still developing. Of course, there are difficulties and challenges, and more experiments are still needed for exploration and testing. For example:

(1) **Accurate Dynamic Gesture Recognition.** Hand gesture and motion acquisition technology is still an important technology that restricts free gesture interaction. Especially for the recognition and acquisition of dynamic gestures, the accuracy of interactive gesture motion capture and recognition should be improved.

(2) **Gesture Detection Under Different Light, Color and Other Complex Backgrounds.** The gesture background is simple in most of the existing gesture detection processes, but the actual background in practical applications is more complex. We can work in any complex environment, such as a teacher's teaching scenario. Therefore, it will be an important topic in the future to study how to improve the accuracy of gesture recognition in complex background.

(3) **Delay in Gesture Recognition.** As gesture recognition requires a complex process, different technologies are required at different stages to complete it, and problems in any step will affect the whole gesture recognition process. Therefore, a perfect gesture recognition architecture is particularly important.

4.2. User Experience. Researchers have been trying to improve user experience while carrying out technological innovation. However, there are still problems to be improved and solved.

(1) **Gesture Learning.** Different gesture effects should be achieved in different task scenarios. Some gestures require users' additional learning and memorizing. Too many or too complex gestures will increase the burden of users' memory.

(2) **Transformation from Traditional Interaction Form to Midair Gesture Interaction.** It is a gradual process of the transition from traditional mouse and keyboard to midair gesture. Although it tends to be difficult for users to get rid of their previous usage habits, midair gesture interaction is not completely divorced from the original form of interaction. Designers need to find a balance between the two to improve user satisfaction.

(3) **Midair Gesture and Multimodal Interaction.** In multimodal interaction scenarios, each interaction mode has its own unique role. There is no agreement on which part of the interface functions are more suitable for gesture manipulation or the combination of gesture and speech, which is particularly important for improving the user experience.

(4) **Perception of Interactive Information.** The research of AR interaction focuses on how users interact with virtual objects. Therefore, the problems to be solved in gesture interface design are how users perceive such interaction, which information can be or cannot be interacted with, which information needs gesture interaction, and how to give users correct judgment.
Definition of Free Midair Gesture. Researchers of user experience are pursuing a common goal: to achieve free midair gesture interaction. However, a consensus is to be reached on what kind of gesture can be called free gesture interaction.

5. Future of Midair Gesture Interaction

5.1. Prospect of Midair Gesture in Education, Health Care, Industry, and Other Professional Fields. The application field of midair gesture interaction is expanding. Depending on continuous breakthrough, AR interaction technology will produce more practice value regarding its application in education, health care, industry, office, and other scenarios. For example, in medical, surgeons use make gestures in front of the camera using computer vision technology to achieve operations such as zooming, rotating, image clipping, and switching slides, which avoids repeated disinfection when doctors use other equipment. In addition, the application also covers surgical training, psychotherapy, etc. The reform of medical education will occur with immersive learning tools built for medical students and nursing professionals. Various 2D and 3D data (such as X-ray, ultrasound, and human structure) are displayed in AR to help surgeons practice the surgical process. Gesture recognition could also create a better life for some disabled patients. Some researchers are also exploring how to transform the sign language of deaf mutes into written language by gesture recognition, for communicating with others in an environment combining virtuality and reality.

In addition, the way of education has also experienced great changes under the influence of normalized COVID-19 control and prevention. And distance teaching is also a great challenge for teachers. In addition to AR experiments, the directions of future exploration are drawing demonstrations, teaching demonstrations in STEM education, and Gamification course content presentations. At present, the exploration of midair gestures in education is still in the development stage. Also, there are many pain points of interactive experience: poor user experience of the technology and lack of real feedback and emotional interaction, resulting in unsatisfactory educational results. Therefore, an important aspect of realizing natural gesture interaction will be designing a more natural way of gesture interaction, so that the boring and profound knowledge can be conveyed to students in a more vivid form without affecting the teacher’s experience or causing fatigue. As shown in Figure 6.

5.2. Cooperative Gestures to Improve Work Efficiency. In many jobs, communication, cooperation, and sharing between people are essential. A more immersive interactive experience will be brought by multiple users operating virtual content through gestures and working together to share virtual space. Most of the current application research of midair gesture only tracks a single framework and supports the experience of a single user. The interactive effect of collaborative presentation can be enriched with the merging of multiple users. For example, Saquib et al. [24] thought that the sense of experience and interest of interaction with the audience can be enhanced, and more diverse cooperative actions can be created, if they can track multiple skeletons and realize the performance of multiperson cooperation.

Remote collaboration and remote communication are the future ways to improve work efficiency; no matter it is in study, office, or industrial production. Especially in the face of normalized COVID-19 control and prevention, AR makes remote work more efficient and convenient. In the workspace of remote collaboration, collaborative interaction has become an essential part. Also, collaborative gestures will play a greater role to help improve work efficiency. AR is shifting towards a social model, and collaborative gestures are also worth exploring in the future.

The metaverse world will also be a world of interoperability and social integration. The future metaverse space will cover a variety of devices, and gesture interaction will be an important part to improve the sense of interactive experience in such virtual experience layer based on physical space. Currently, speech communication still dominates in the metaverse space, and it is difficult to realize the simulation of sensory and body language. Besides, it does not have the sense of social space and cannot provide a fully immersive
experience. But, all this will be the direction of future efforts.
Metaverse space is making natural and immersive AR interaction experience a reality with the continuous innovation of technology.

5.3. Low-Threshold and Low-Cost User Experience. The experience of augmented reality nowadays still depends on high-cost hardware devices, such as AR glasses, which often discourages many users and also limits the diversified development of gesture interaction. Therefore, more low-threshold and low-cost entrances (e.g. smart phones) should be opened up to the world integrating virtuality and reality. According to the third report of the long-term survey of 《AR Usage and Consumer Attitudes Report of ARtillery 》 in 2020, it is expected that the penetration rate of mobile AR users will almost coincide with that of Internet users by 2023, which means that all Internet users may be mobile AR users. With smart phones as the entrance, gestures are undoubtedly the fastest way of interaction. Mobile phone users can achieve very interesting results by painting and placing mobile AR objects with midair gestures anytime and anywhere. Especially in the field of teaching, the previous mobile AR is more about the enhancement of visual effects, and the transformation of 2D plane effects into 3D stereo vision [38] (Figure 7). If one can also interact with it with midair gestures, a more immersive experience will be produced, with the advantages of saving time and cost. Figure 8 shows the design and implementation case of AR midair gestures interaction of art works in art exhibition by Weng et al. [39].

At present, mobile AR is facing many problems. In mobile AR, users have to raise their arms for a long time, with one hand for gesture interaction. In consequence, inaccurate AR positioning and gesture recognition will occur, resulting in poor AR experience and shortened user experience time. Therefore, it is necessary to consider enhancing user utilization by both enhancing the user experience and improving technology. It is believed that accurate midair gesture interaction with virtual objects can be realized on mobile phones with the development of technology.

6. Summary and Discussion
With a focus the midair gesture interaction in AR, the literature and main results in the field of midair gesture interaction were systemically studied by the literature research method. From the perspective of user experience, the gesture types in midair gesture interaction and the typical applications of midair gesture in AR as well as the challenges and future of midair gesture interaction were analyzed. The aim is to give reference and interactive experience design suggestions to designers in gesture interaction research.
Midair gesture interaction is a more natural and flexible interaction method, and will be an important interaction method in AR. In terms of the application of midair gestures, a detailed analysis was carried out on the application and potential research value of midair gestures in the field of education. In the field of education, air gesture interaction for teaching has important research value because of the particularity of experience environment, user type, and interaction purpose. The user experience of gesture interaction needs to be considered from the educational attribute, functional attribute, and interactive attribute. In addition, the combination of gestures and multisense and multichannel can also meet the user needs for immersive experience.

There are still challenges in gesture recognition technology and user experience. The earliest gesture classification comes from communication, while the classification of midair gestures in AR will be detailed to specific task types and lay more emphasis on user experience. In most of the existing research results, the evaluation of gestures is implemented in the laboratory environment. Because of the diversity of gestures and the universality of task types, it is difficult to form a general gesture set, which requires researchers to design corresponding gesture types for different task needs and carry out multiple experiments in real scenes, rather than just laboratory research. For a wider application, more field evaluations are needed to understand what gestures are effective for what tasks, so as to provide a better user experience. Midair gesture interaction is a free input method, but for the user experience, it is not only necessary to consider the freedom of the hand, but also to consider the overall comfort from the perspective of ergonomics. At present, most gesture output devices in AR are still head mounted displays, which are not suitable for long-term wearing. Therefore, the hardware design also needs to be fully evaluated. In addition, from the application cases of midair gestures, we can see that the use types of gestures are not complex, such as rotation, translation, and switching, which are related to two factors; one is limited by gesture recognition technology. Gesture recognition is a complex process. To achieve accurate gesture recognition, a perfect technical framework is required. Therefore, midair gestures are not suitable for overly accurate interactive tasks; on the other hand, simple gestures can reduce the burden of users’ learning and memory, which is more suitable for user experience. It can be said that it is a trend of interaction that simple gestures control diverse content.

In a word, designers need to consider all aspects of technology, experience, and innovative applications to enhance user stickiness and personalized experience. With the continuous development of AI, hybrid reality and gesture recognition technology, midair gesture interaction will be more natural and comfortable, and its application field will be more extensive.

**Data Availability**

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Conflicts of Interest**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**References**

[1] R. Azuma, “A survey of augmented reality,” Presence: teleoperators & virtual environments, vol. 6, no. 4, pp. 355–385, 1997.
[2] Y. Wu, X. Zhang, and P. W. Fu, Human Computer Interaction Technology and Application, China Machine Press, 2020.
[3] T. Vuletic, A. Duffy, L. Hay, C. McTeague, G. Campbell, and M. Grealy, “Systematic literature review of hand gestures used in human computer interaction interfaces,” International Journal of Human-Computer Studies, vol. 129, pp. 74–94, 2019.
[4] Z. Wei, L. Zeyi, C. Jian, K. Mingyu, D. Xiaoming, and W. Hong’an, “Overview of dynamic gesture understanding and interaction,” Journal of software, vol. 32, no. 10, pp. 3051–3067, 2021.
[5] C. Groeneveld, C. Anslow, J. Islam, C. Rooney, P. J. Passmore, and B. L. Wong, “Understanding 3D mid-air hand gestures with interactive surfaces and displays: a systematic literature review,” in Proceedings of the 30th International BCS Human Computer Interaction Conference (HCI), Poole, United Kingdom, 2016.
[6] A. Kendon, “How gestures can become like words,” in This Paper Is a Revision of a Paper Presented to the American Anthropological Association, Chicago, Dec 1983, Hogrefe & Huber Publishers, 1988.
[7] C. Cadoz, Les Réalités Virtuelles, Flammarion, Paris, 1994.
[8] D. McNeill, Hand and Mind: What Gestures Reveal about Thought, The University of Chicago Press, Chicago, London, 1995.
[9] M. Karam and M. C. Schraefel, A taxonomy of gestures in human computer interactions. Technical report ECSTRAIMOS-009, electronics and computer science, University of Southampton, 2005.
[10] R. Aigner, D. Wigdor, H. Benko et al., “Understanding mid-air hand gestures: a study of human preferences in usage of gesture types for hci,” vol. 2, Microsoft Research Tech Report MSR-TR-2012-111, 2012.
[11] T. Wang, X. Qian, F. He, X. Hu, Y. Cao, and K. Ramani, “GestureAR: an authoring system for creating freehand interactive augmented reality applications,” in The 34th Annual ACM Symposium on User Interface Software and Technology, pp. 552–567, Virtual Event, USA, 2021.
[12] M. Moran-Ledesma, O. Schneider, and M. Hancock, “User-defined gestures with physical props in virtual reality,” Proceedings of the ACM on Human-Computer Interaction, vol. 5, no. ISS, pp. 1–23, 2021.
[13] N. Samimi, S. von der Au, F. Weidner, and W. Broll, “AR in TV: design and evaluation of mid-air gestures for moderators to control augmented reality applications in TV,” in 20th International Conference on Mobile and Ubiquitous Multimedia, pp. 137–147, Leuven, Belgium, 2021.
[14] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, “User-defined gestures for surface computing,” in Proceedings of the 27th international conference on Human factors in computing.
systems - CHI 09. Volume 8, p. 1083, New York, New York, USA, 2009.

[15] M. Lee, "Multimodal speech-gesture interaction with 3D objects in augmented reality environments (2010)," in Department of Computer Science and Software Engineering, University of Canterbury, Christchurch, 2010.

[16] T. Piumsomboon, A. Clark, M. Billinghamurst, and A. Cockburn, "User-defined gestures for augmented reality," in IFIP Conference on Human-Computer Interaction, pp. 282–299, Springer, Berlin, Heidelberg, 2013.

[17] G. A. Rovelto Ruiz, D. Vanacken, K. Luyten, F. Abad, and E. Camahort, "Multi-viewer gesture-based interaction for omni-directional video," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 4077–4086, Toronto, ON, Canada, May 2014.

[18] S. Siddhpuria, K. Katsuragawa, J. R. Wallace, and E. Lank, "Exploring at-your-side gestural interaction for ubiquitous environments," in Proceedings of the 2017 Conference on Designing Interactive Systems, pp. 1111–1122, Edinburgh, UK, June 2017.

[19] I. A. Zaiti, S. G. Pentiuc, and R. D. Vatavu, "On free-hand TV control: experimental results on user-elicited gestures with leap motion," Personal and Ubiquitous Computing, vol. 19, no. 5-6, pp. 821–838, 2015.

[20] N. K. Dim, C. Silpasuwanchai, S. Sarcar, and X. Ren, "Designing mid-air TV gestures for blind people using user- and choice-based elicitation approaches," in Proceedings of the 2016 ACM Conference on Designing Interactive Systems, pp. 204–214, Brisbane, Australia, June 2016.

[21] H. Rateau, L. Grisoni, and B. De Araujo, "Mimetic interaction spaces: controlling distant displays in pervasive environments," in Proceedings of the 19th International Conference on Intelligent User Interfaces, pp. 89–94, Haifa, Israel, February 2014.

[22] K. Perlin, Z. He, and F. Zhu, "Chalktalk VR/AR," Artificial Intelligence Meets Virtual And Augmented Worlds, 2010, No 2017/2 (2017): Proceedings of the 10th International Workshop on Semantic Ambient Media Experiences (SAME 2017): Artificial Intelligence MEETS Virtual and Augmented Worlds (AVIR) - in conjunction with SIGGRAPH Asia.

[23] J. H. Lee, D. Ma, H. Cho, and S. H. Bae, "Post-post-it: a spatial ideation system in VR for overcoming limitations of physical post-it notes," in 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–7, Yokohama, Japan, 2021.

[24] N. Saquib, R. H. Kazi, L. Y. Wei, and W. Li, "Interactive body-driven graphics for augmented video performance," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12, Glasgow, Scotland UK, 2019.

[25] J. Gong, T. Han, S. Guo et al., "Holo board: a large-format immersive teaching board based on pseudo holo graphics," in The 34th Annual ACM Symposium on User Interface Software and Technology, pp. 441–456, Virtual Event, USA, 2021.

[26] L. Geping and N. Gao, "The influence mechanism of gesture interactive virtual experiment on learning experience," Research on Modern Distance Education, vol. 33, no. 2, pp. 22–32, 2021.

[27] P. Meenakshi and S. M. Pawan, "Hand gesture recognition for human computer interaction," in 2011 International Conference on Image Information Processing, pp. 1–7, Shimla, India, 2011.

[28] C. Ananya, T. Anjan Kumar, and K. S. Kandarpa, A Review on Vision-Based Hand Gesture Recognition and Applications, vol. 11, IGI Global, 2015.

[29] M. Yasen and S. Jusoh, "A systematic review on hand gesture recognition techniques, challenges and applications," Peer J Computer Science, vol. 5, article e218, 2019.

[30] N. S. Saehee and A. W. Ismail, “AR home deco: virtual object manipulation technique using hand gesture in augmented reality,” Innovations in Computing Technology and Applications, vol. 3, 2018.

[31] J. Y. Lee, G. W. Rhee, and D. W. Seo, "Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment," International Journal of Advanced Manufacturing Technology, vol. 51, no. 9-12, pp. 1069–1082, 2010.

[32] T. Lee and T. Hollerer, "Handy AR: markerless inspection of augmented reality objects using fingertip tracking," in 11th IEEE International Symposium on Wearable Computers, ISWC 2007, pp. 83–90, Boston, MA, USA, October 2007.

[33] T. Lee and T. Hollerer, "Multithreaded hybrid feature tracking for markerless augmented reality," IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 3, pp. 355–368, 2009.

[34] B. Fernandes and J. Fernandez, "Bare hand interaction in tabletop augmented reality," in SIGGRAPH 2009: Posters, Association for Computing Machinery, 2009.

[35] V. Buchmann, S. Violich, M. Billinghamurst, and A. Cockburn, "Fing ARTips: gesture based direct manipulation in augmented reality," in Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia, pp. 212–221, Singapore, 2004.

[36] H. Benko, R. Jota, and A. Wilson, "Miragetable: freehand interaction on a projected augmented reality tabletop," in Proceedings of the SIGCHI conference on human factors in computing systems, pp. 199–208, Austin Texas, USA, 2012.

[37] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson, "Holo desk: direct 3D interactions with a situated see-through display," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin Texas, USA, 2012.

[38] M. C. Hsieh and J. J. Lee, "Preliminary study of VR and AR applications in medical and healthcare education," Journal of Nursing and Health Studies, vol. 3, no. 1, p. 1, 2018.

[39] C. Y. Weng, B. Curless, and I. Kemelmacher-Shlizerman, “Photo wake-up: 3D character animation from a single photo,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5908–5917, Long Beach, CA, USA, 2019.