Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/42/20230778

Multimodal human-computer interaction for virtual reality

Jiajun Tang[1,6,†], Manni Gong[2,†], Shangxi Jiang[3,†], Yiji Dong[4,†], Taojie Gao[5,†]

[1]Shool of Advanced Technology, Xi'an Jiaotong-liverpoor University, Suzhou, China
[2]Wuhan Britain-China School, Wuhan, China
[3]Suzhou Science &Technology Town Foreign Language School, Suzhou, China
[4]Jinan Foreign Language School, Jinan, China
[5]Wuhan Britain- China School, Wuhan, China

[6]Corresponding author: Jiajun.tang21@student.xjtlu.edu.cn
[†]Jiajun Tang, Yiji Dong, Taojie Gao, Manni Gong and Shangxi Jiang contributed equally to this work and should be considered co-first authors.

Abstract. Virtual reality (VR) is an interactive simulation environment based on computer technology that enables users to immerse themselves in a virtual world. With the continuous advancement of technology, VR has shown great potential in various fields. This article aims to review the application of VR in different fields and its impact on society. First, this paper introduces VR's basic principles and technical characteristics. VR systems are usually composed of head-mounted display devices, trackers, joysticks, etc., enabling users to experience immersive virtual environments by simulating real-world perception and interaction. Secondly, this paper discusses the application of VR in the field of education. VR technology can provide an immersive learning experience, enabling students to personally participate in the course content, greatly enhancing the effect of learning. At the same time, VR can also simulate experimental scenes, train operational skills, etc., to provide students with more abundant and practical learning resources. Then, this paper introduces the application of VR in the medical field. VR can be used for rehabilitation training of patients, psychological treatment, and other aspects, effectively improving the quality and effect of medical services. In addition, this paper also discusses the application of VR in entertainment, tourism, and other fields. VR technology can bring users immersive gaming experience, virtual tourism, etc., and inject new vitality into the entertainment industry. In tourism, VR can provide tourists with an immersive sightseeing experience, so they can feel the distant scenery at home.

Keywords: multimodal interaction, virtual reality, styling, human computer interaction, eye tracking

1. Introduction

Virtual Reality (VR) is a kind of simulation environment simulated by computer technology, allowing users to immerse themselves in a virtual three-dimensional space and interact with the environment. VR technology typically includes hardware devices such as head-mounted displays, tracking devices and interactive devices, and virtual reality software and applications.

The core goal of virtual reality technology is to create an immersive feeling, making users feel like they are in a virtual world. Through the head-mounted display, the user can see a panoramic virtual environment, while the tracking device can track the user's head and hand movements, allowing the user to freely move and interact in the virtual environment. Interactive devices such as joysticks and gloves allow users to interact with objects in a virtual environment [1, 2].

The development of VR technology has been widely used in many fields. In the field of gaming, VR technology provides players with a more immersive gaming experience, allowing them to participate in the game world. In the field of entertainment, VR technology can be used to create experiences such as virtual cinemas and virtual Tours, allowing users to feel different scenes and situations. In the field of education, VR technology can be used to create virtual LABS, virtual classrooms, etc., to provide a more vivid and interactive way of learning. In the medical field, VR technology can be used for surgical simulation, rehabilitation training, etc., to help doctors and patients' better treatment and rehabilitation.

The development of virtual reality technology cannot be separated from the support of hardware equipment and software applications. A variety of VR headsets are available on the market, such as Oculus Rift, HTC Vive, Sony PlayStation VR, and others. These devices all have high-resolution displays and accurate tracking systems that provide a more realistic virtual experience. At the same time, many software and applications support the development and application of VR technology, such as Unity, Unreal Engine and other game engines, as well as various virtual reality applications.

The development prospect of virtual reality technology is broad but also faces some challenges. One of them is the cost and performance of hardware devices, and current VR devices are expensive and not yet popular enough for the average consumer. In addition, virtual reality technology also needs more content and application support to meet users' needs.

In short, virtual reality technology allows users to be immersive and has been widely used in games, entertainment, education, medical and other fields. With the continuous development of hardware devices and software applications, virtual reality technology is expected to bring people a more realistic and immersive experience.

2. The Working Principle of VR

2.1. Specific operating principles

VR technology strives to construct immersive virtual experiences by engaging multiple sensory modalities. Virtual reality technology mobilizes multiple sensory modalities, including vision, hearing, and touch. Vision is presented through a head-mounted display (HMD), built-in spatial audio gives auditory realism, and haptic feedback simulates contact with physical objects. The process begins with VR equipment capturing participants' physical attributes, such as height and hand positions. When a user moves, there is a need for accurate tracking of the user's movements. Generally built-in sensors, cameras, and motion trackers monitor changes in the user's head, hand, and body postures to ensure stability and safety of movement. Subsequently, this bodily data is integrated into the device and channeled through specialized software offered by the VR device provider.

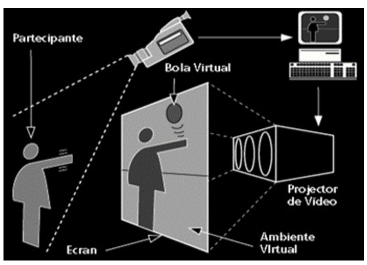

Figure 1. Working Principles of VR technology. [3]

This software orchestrates the substitution of the user's surroundings with a meticulously crafted virtual environment (Fig. 1). Once this substitution is achieved, the scene undergoes rendering. It requires real-time rendering of high-quality graphics to ensure stable interactions and prevent

DOI: 10.54254/2755-2721/42/20230778

discomfort. This usually requires a device with a powerful graphics processor (GPU) to render and display at a high frame rate, minimizing the effects of latency. After rendering completion, the output is showcased through a head-mounted display. The culmination of this process materializes as users don the display apparatus, enabling them to explore the synthetic reality uninhibitedly within their actual physical surroundings

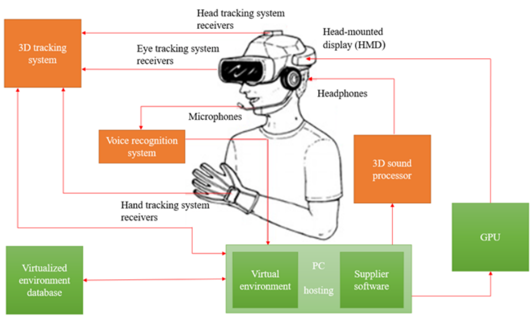

Figure 2. Principle of tracking system (original)

2.2. *Multimodal feedback*

The human perception of the world is a real-time, intricate process facilitated by the nervous system's integration of various sensory modalities [4]. Perception is how our brain combines information from our senses like smell, taste, sight, hearing, and touch, helping us make sense of the world around us and adapt to it. Notably, virtual reality (VR) technology facilitates multi-modal human-computer interactions. Achieved through head-mounted display devices outfitted with adjustable motors and an array of tracking technologies, these systems effectively replicate key human perceptions, encompassing visuals, audio, touch, and more. H. Xun et al. [5] emphasize that current interaction techniques for VR head-mounted displays predominantly encompass sound control, gesture recognition, head motion tracking, and eye tracking. It is acknowledged, however, that voice and gesture control face limitations within certain environments and ranges of activity, thereby relegating head motion tracking as the prevailing mode of interaction. This reliance raises concerns as frequent neck movement, given the head's weight of approximately 5kg (7% of total body weight), can potentially lead to vertigo and cervical spine injuries. Conversely, the eyes, each weighing around 8g (a mere 0.002% of body weight), offer a more adaptable and less burdensome means of movement. Moreover, the intricate ocular muscles responsible for eye motion exhibit resistance to fatigue and physical strain, rendering eye-tracking technology both swift in response and remarkably precise in recognition. Perception is how our brain combines information from our senses like smell, taste, sight, hearing, and touch, helping us make sense of the world around us and adapt to it.

Eye tracking is a technology that monitors the user's gaze and eye movements. It has the significant advantage of measuring the user's focus in a scene in real time and recording the behavior of their visual system. Currently there are three dominant eye tracking techniques are video-based ophthalmoscopy (VOG) video-based infrared (IR) pupil-corneal reflection (PCR), and (3) electrooculography (EOG), of which the most commonly used technique is VOG. which is specialized video-based tracking widely used in head-mounted displays. A typical system consists of a camera that records eye movements and a computer that saves and analyzes gaze data. In a head-mounted system, the camera is either mounted on the frame of the eyeglasses or in a separate "helmet." Head-mounted systems also typically include a scene camera for recording the user's perspective, which can then be used to map the user's vision to the current visual scene. In a head-mounted system, the camera is placed near the eyes, which means that images of the eyes are better acquired, and therefore the pixels tracking the eyes are higher. Better results are obtained if a wide angle camera is used, but this means that the user will have more freedom of movement, and a higher resolution camera will be needed to maintain adequate pupil tracking accuracy. Since the tracking is based on video images of the eyes, an

DOI: 10.54254/2755-2721/42/20230778

unobstructed view of the eyes is required. Various factors also affect the quality of tracking, such as lighting conditions, spectacle reflectivity, eyelid droop ratio, eye condition when smiling, and even heavy makeup [6]. In addition, VOG Estimation of gaze position by pupil tracking and corneal reflection analysis involves capturing the light emitted by the eye. As a result, when visual attention is diverted away from the designated area of focus, the accuracy of the eye tracker decreases, highlighting a key limitation of eye tracking [7].

3. The Defects and Development Prospects of Human-Computer Interaction

3.1. Flaws in human-computer interaction

Although virtual reality technology is developing rapidly, some constraints exist between its human-computer interaction. First, it is difficult to recognize information. For most users, it is difficult to grasp the normal flow of a VR device, which means that if they can't follow a mechanically fixed process, they will inevitably deal with frequent errors, which directly affects the effectiveness of human-computer interaction and causes the patience and interest of the experiencer to drop precipitously. For example, the mainstream intelligent voice systems on the market, Apple's Siri, Microsoft's Cortana, Google's Android's Google Now, etc., these systems in the process of human-computer interaction are generally more mechanical and dull, and cannot lead to make the user desire to interact with them, and, unfortunately, the majority of users will only use them to entertain themselves when they are relatively bored, and will not expect them to be able to solve their own problems. So this is going to be one of the most important criteria for weighing the good and bad of VR human-computer interaction.

Meanwhile, from a technical point of view, the interference and recognition of voice interaction is relatively complex. Dialects, colloquialisms, and slang may affect automatic speech recognition (ASR) results. Voice interaction, the requirements for intelligent speech recognition system are higher, because it is the transmission bridge and link between human and device interaction. And with so many languages and emotions, it's hard to recognize exactly what the user wants to say, and the accuracy is much lower

At this time, there are some new attempts in human-computer interaction technology, such as somatosensory interaction, eye tracking, voice interaction, biometrics, etc., but the real use of these technologies for real use is not high. In other words, the limitation of the scope of human-computer interaction makes it has not yet entered the stage of commercial application popularization in the real sense, and it cannot meet the idealized human-computer interaction to achieve the degree of barrier-free and casual communication between human and equipment (machine). In addition to a number of devices near the use of somatosensory interaction to operate will have certain disadvantages, for example, in the field of gaming will have limitations, in the game environment within the faults and errors will not cause too serious damage, and the use of virtual reality technology will be directly linked to the reality, perhaps through the head of the monitor shows the environment is safe, but in the process of playing due to the constraints of the site will be the real-life, it is possible to run into the Around the sharp objects or knocked into the surrounding walls, if the use of somatosensory technology for human-computer interaction, cannot be correctly combined with the real environment will affect the personal safety of the experiencer. In addition to the technical limitations received, relevant experts' research found that the full use of immersive VR equipment for too long will produce motion sickness. So the use of virtual reality technology for human-computer interaction so far still has certain disadvantages, but also need long-term research and development. [6]

3.2. Development prospect

As an important branch of computer science, human-computer interaction (HCI) is constantly evolving with the push of technology. Virtual reality (VR) technology enables users to interact with the virtual world by creating a fully immersive virtual environment. In the future, virtual reality technology will be more realistic and fluid, e.g., providing a more immersive virtual experience through higher-resolution head-mounted displays, more accurate tracking techniques, and more natural interactions. A novel hand-adaptive user interface technique is proposed in the paper "Improving Hand Gesture Interaction in Virtual Reality" by X.L. Lou et al. [7] to improve hand interaction performance in VR. Compared to traditional user interfaces, hand adaptive user interfaces offer higher interaction efficiency, lower physical effort and perceived task difficulty. Utilizing this technology, more convenient human-computer interaction products will be developed.

Meanwhile, virtual reality (VR) will likely play a transformative role in the field of HCI learning. Following the reasoning of M. Lui et al. [8] who expanded the space of possibilities for embodied learning in VR by experimenting with undergraduate students using simulated scientific models in head-mounted displays to address a wider range of difficult-to-teach abstract concepts. Because of the immersive learning environments that can be created with the help of virtual reality, the use of human-computer interaction will focus more on breakthroughs in perceptual technologies, including sight, sound and touch. Natural language processing and intelligent assistants will give students a more natural and smooth dialog experience. This will lead to increased immersion and interaction. Students can explore historical events, visit distant places, and even travel inside the human body, making abstract concepts more visual and engaging. So in the future, VR human-computer interaction will be used more in education, healthcare and other fields, bringing greater convenience and efficiency to people's life and work.

4. Application

When the principle of human-computer interaction becomes clearer and more specific, how to apply these technologies to real life become the main concern. Taking advantages of collaboration of human and computer collaboration can improve safety and efficiency through many tasks, and even can fulfill part of job-vacancy. For example, in the field of medical care, there are currently some wheelchairs on the market that use virtual reality technology to better help patients.

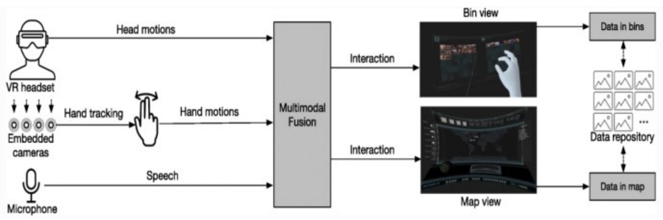

Figure 3. Multimodal real-time interaction for VR interfaces [9]

4.1. Application to Smart Wheelchairs

The report by R. S. Rao et al. [10] demonstrates their research on human-computer interaction in smart wheelchairs, where they have been able to solve the problem of autonomous use of wheelchairs by patients with hearing and speech impairments through the implementation of a smart wheelchair interaction interface using visual interaction. Different levels of personalization are provided for the user to choose. During the use of the wheelchair, the user receives a series of images that are monitored by an overhead camera through projection, and the images they receive are converted to projections. The technology uses virtual reality to ensure that a virtual keyboard and monitor can be used throughout the system, rather than the less accessible moving mechanical parts of the wheelchair. This makes the user aware of obstacles around the wheelchair, which encourages a smoother, more continuous operating experience. Interaction commands and image analysis are then sent to the processor, where algorithms will call back interaction responses.

The whole system is very user-friendly as it is designed to meet the user's different needs. In detail, depending on the analyzed data, the wheelchair allows the simplest direct access. Additionally, the controller can respond to a more complex set of motor behaviors. In addition, the top level allows the wheelchair to automatically navigate to the desired destination, and route planning also prevents rough terrain to avoid inconvenience. Besides, the operating system will provide emotion icons for disabled people to communicate, such as thank you, apology and applause, and the emotion icons also have corresponding voice packs for communication. Ultimately their system utilizes the construction of a virtual interactive interface to provide powerful assistance to people with disabilities.

4.2. Application to Tourism

With the popularity of virtual reality, the application of virtual reality is used in many fields. In addition, virtual reality provides a new platform for visitors to communicate and exchange information that can enhance their experience.

Many legacy sites and artifacts worldwide have been digitized into 3D virtual models that can be overlaid on top of each other to create realistic and navigable virtual environments that can be explored from different perspectives, going beyond the traditional 2D blueprints and static 3D models. But many of these are not available to the public, as it has the added incomparable advantage that the use of virtual reality provides a powerful means of visualizing and experiencing spatial environments, making it an important tool for urban, environmental and architectural planning The list of heritage sites and objects that can be viewed electronically is constantly growing. In addition to those already mentioned, there are many other examples of historical sites and objects that have been reproduced as 3D models Utilizing virtual tours, the use of a visual language promotes the exchange and cooperation of cultures from around the world, thus facilitating the participation of local communities in the tourism planning process. Because such virtual models can include incredibly precise and accurate data sets that can theoretically be maintained in perpetuity, they can serve as an important tool for cultural preservation. Such locations and items can be rendered as virtual 3D models. It is sometimes necessary to compensate for centuries of natural degradation or destruction through heritage conservation and restoration activities. However, some heritage sites require conservation efforts, such as the use of virtual reality that can simulate environmental impacts such as soil erosion, contributing to sustainable land-use planning in tourist destinations so that they can continue to be developed and appreciated in the future as tourist attractions [11].

5. Conclusion

Human-Computer Interaction (HCI) refers to the technology that realizes the dialogue between human beings and computers in an effective way through computer input and output devices. In this paper, we provide an overview of the field of human-computer interaction in terms of its principles, applications, limitations, and future developments. The main findings of the paper are as follows:

Nowadays, the most human-computer interaction that people come into contact with in their lives is divided into two forms: head-mounted and wearable. Head-mounted is often referred to as VR, in which the simplest form of interaction is timed aiming. A tablet and a back button are added in more advanced mobile VR devices, such as the Samsung Gear VR. Furthermore, wearable device is commonly understood as a wearable portable computing device with miniaturization, portability, small size and strong mobility. The core technology of the device has bone conduction interaction technology, eye tracking interaction technology, voice interaction technology and other seven interaction technologies.

However, HCI technology has a series of problems in the present day. The first is that there is still a lot of room for improvement regarding the accuracy and real-time performance of natural perception technology, so the information recognition of this technology is difficult. In addition, human-computer interaction in the current application field is mainly in the medical field, for other fields are not completely popular, which is why many people still do not understand this technology. There are also risks associated with applying HCI to the field of driving technology.

In this paper, we also list applications of human-computer interaction.

Wheelchair utilizes a combination of vision-based HCI, a set of sensors and an intelligent control algorithm that allows the user to select different levels of automation. Tourism is also an application. With the popularity of virtual reality in the tourism industry, many scenic spots attract more tourists through innovative tours. In addition, virtual reality also provides a new information communication

DOI: 10.54254/2755-2721/42/20230778

and exchange platform for tourists, which enhances their experience. Finally, it is entertainment. Interactive film is also an application of virtual reality, in which the plot of the film will change according to the visual interests of the audience.

Acknowledgement
Jiajun Tang, Yiji Dong, Taojie Gao, Manni Gong and Shangxi Jiang contributed equally to this work and should be considered co-first authors.

References

[1] B. Federick.” Myron Krueger.” PHWORKS.com. http://thedigitalage.pbworks.com/w/page/22039083/Myron%20Krueger (accessed Nov. 23, 2009).
[2] E. Bozkir, “Towards Everyday Virtual Reality through Eye Tracking,” arXiv.org, Mar. 2022.
[3] Han et al,” Patent analysis of augmented reality/virtual reality on head mounted display,” Hebei Journal of Industrial Science & Technology, vol. 34, no. 5, pp. 345–353, 2017.
[4] P. Majaranta and A. Bulling, Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies. IGI Global, 2011.
[5] A. D. Souchet, S. Philippe, D. Lourdeaux, and L. Leroy, “Measuring Visual Fatigue and Cognitive Load via Eye Tracking while Learning with Virtual Reality Head-Mounted Displays: A Review,” International Journal of Human–Computer Interaction, vol. 38, no. 9, pp. 1–24, Sep. 2021.
[6] S. Stadler, H. Cornet, and F. Frenkler, “Assessing Heuristic Evaluation in Immersive Virtual Reality—A Case Study on Future Guidance Systems,” Multimodal Technologies and Interaction, vol. 7, no. 2, pp. 19, Feb. 2023.
[7] X. Lou, X. A. Li, P. Hansen, and P. Du, “Hand-adaptive user interface: improved gestural interaction in virtual reality,” Virtual Reality, vol. 25, no. 2, pp. 367–382, Jul. 2020.
[8] M. Lui, Kwok Ying Chong, M. Mullally, and R. McEwen, “Facilitated model-based reasoning in immersive virtual reality: Meaning-making and embodied interactions with dynamic processes,” International Journal of Computer-supported Collaborative Learning, vol. 18, no. 2, pp.203-230, Aug. 2023.
[9] L.Cao, H. Zhang, C. Peng, and J. T. Hansberger, “Real-time multimodal interaction in virtual reality - a case study with a large virtual interface,” Multimedia Tools and Applications, vol. 82, no. 16, pp. 25427–25448, Feb. 2023.
[10] R. Rao, T. Conn, S. Jung, J. Katupitiya, T. Kientz, V. Kumar, J. Ostrowski, S. Patel, C. Taylor, “Human robot interaction: application to smart wheelchairs,” Scholarly Commons (University of Pennsylvania), Jun. 2003.
[11] D. Guttentag, “Virtual reality: Applications and implications for tourism,” Tourism Management, vol. 31, no. 5, pp. 637–651, Oct. 2010.