Human-computer Interaction in Augmented Reality

Weihua Pan [a], Zhihao Jin [b]

School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China.

[a]panwh@ncepubd.edu.cn, [b]iqinfinity@qq.com

Abstract

Human-computer interaction is one of the core technologies of augmented reality, and plays an important role in promoting the construction of augmented reality and improving the user experience. With the development of sensors and other hardware technologies, human-computer interaction in augmented reality has also made great progress. This paper mainly introduces the human-computer interaction technology and equipment in augmented reality, and analyzes its development trend.

Keywords

Augmented reality, human-computer interaction, user interface, 3D, multi-channel.

1. Introduction

Human-computer interaction is a multidisciplinary field of computer, ergonomics, engineering psychology, and cognitive science. Human-computer interaction was first proposed in 1975. The professional name appeared in 1980, and in 1983, Card, Moran, and Vail published the book Human-Computer Interaction Psychology, and the concept of interaction has spread rapidly since then. Human-computer interaction research was originally machine-centric, and psychologists trained and selected employees to adapt to the machine. Later, during the Second World War, the machine was so complicated that it was difficult to adapt people. The focus of research shifted to the human-centered, and how the machine adapts to the psychological characteristics of people. From the emergence of the first computer to the arrival of the Internet era, the interaction between people and computers has gone through the following stages: early manual operation phase; job control language and command interactive language phase; graphical user interface phase; network user interface phase; multi-channel The intelligent human-computer interaction stage of multimedia. The Internet has penetrated into all walks of life and entered the Internet of Things era. It needs to meet the different needs of different users, especially those with less experience in using electronic devices. The natural user interface is designed according to the psychological mode of people's daily behaviors. People with less experience in using the device interact with the computer as if they interact with the real environment. They do not need to memorize the relevant program functions, which can greatly reduce the user's Memory burden, which can solve the problem that the interface function is complicated and difficult to understand and the related program knowledge is weak. At present, human-computer interaction is developing in the direction of anthropomorphization, intelligence, naturalization and materialization[1].

2. Interface Paradigm and Interactive Tasks

2.1 Interface Paradigm of Augmented Reality

The dominant graphical user interface in the desktop environment is based on the WIMP paradigm that supports the current generation of applications including web browsing, document editing, spreadsheets, application mapping and table games. It provides a fast, accurate and stable way of interacting with the user. However, for applications in augmented reality environments, this kind of interactive organization is not appropriate. For example, in an augmented reality application, the user usually needs to place an object in any position and in any position in the virtual three-dimensional space. For such a task, a two-dimensional mouse is difficult to perform.

The main problem of the graphical user interface supported by the WIMP paradigm is that the input bandwidth in the interaction is too narrow, and the input side only supports precise point and discrete key signals, forming a sequential conversational interaction, which does not make full use of the operational skills that people have learned in real life. In order to break through this limitation, some scholars proposed the concept of Non-WIMP and Post-WIMP[2], trying to break through the limitations of graphical user interfaces, improve the bandwidth of human-computer interaction, and make the interaction process more natural. Since the 1990s, the Post-WIMP interface represented by 3D user interface, hybrid and augmented reality user interface, multi-channel user interface, etc. has quickly become a hot research topic at home and abroad.

In 2006, ACM CHI organized a workshop entitled “What is the Next Generation of Human-Computer Interaction?”, which investigated a large number of interactive technologies with different styles. By the start of the 2008 conference, Jacob proposed a reality-based interaction (RBI)[3], and attempted to guide the interface design as a unified framework for the Post-WIMP interface. The RBI framework proposed by Jacob mainly consists of four levels:

The principle of physics: the universal perception of the physical world. Such as gravity, friction, speed, object presence and scaling;

Human perception and skills: body consciousness, human body's proprioception of the body and the ability to control and coordinate the limbs;

Environmental perception and skills: human perception of the surrounding environment and the ability to operate and navigate in the environment;

Social perception and skills: The ability of humans to perceive others in the environment and interact with others.

The RBI framework provides a guide for interface designers to explore metaphors from real-world behavior to human-computer interaction. Although RBI is not a unified interaction paradigm or model, it provides meaningful guidance for augmented reality and virtual reality user interface paradigm design.

2.2 Interactive Task of Augmented Reality

Augmented reality emphasizes immersion, interactivity and conceiving, which determines that it is different from traditional two-dimensional human-computer interaction. Three-dimensional interaction can enhance the user's perception of the environment and provide natural interaction. It is the most important way to enhance reality. People live in three-dimensional space. They learn a lot of skills in manipulating three-dimensional objects and moving in three-dimensional space around daily life in daily life, and can understand the three-dimensional relationship well. The use of three-dimensional interaction can give full play to the inherent skills of human beings, making augmented reality human-computer interaction more natural and harmonious, and understood by users. Therefore, the three-dimensional interaction oriented to the enhanced environment is the research focus of augmented reality human-computer interaction.

3. Interactive Devices

3.1 Output Devices

Augmented reality systems provide information to one or more sensory organs of the user through output hardware, most of which are primarily used to stimulate human vision, hearing and touch, and in rare cases, to the sense of smell or taste. In augmented reality human-computer interaction, when designing, developing, and using different interaction technologies, careful consideration must be given to selecting display devices, as some interaction technologies may be more applicable than other interaction technologies for a particular display. The augmented reality interactive output device mainly includes a visual output device, an audible output device, and a force/tactile output device.

Visual output device. The visual display transmits information to people through the human visual channel, and is in a very important position in the human-machine system. Augmented reality visual output devices typically include terminal displays, ring screen displays, work bench displays, dome displays, head mounted displays, and cantilever displays.

Audible output device. An important function of the sound display is to enable the participant to utilize his auditory positioning capabilities by generating and displaying spatial three-dimensional sound. Sound is often output as a second form of feedback in augmented reality human-computer interaction, or as an alternative to other sensing channels (such as haptics), which typically use headphones and external speakers to output stereo, surround sound, and 3D audio to the user.

Force/tactile output device. With the corresponding force/tactile devices, people can freely and realistically touch, manipulate and perceive objects in the virtual scene in order to obtain a good sense of realism and immersion[4]. Common force/tactile output devices include data gloves, force feedback mice, force feedback joysticks, force feedback steering wheels, and force feedback arms.

3.2 Input Devices.

An equally important part of augmented reality human-computer interaction design is the selection of a suitable set of input devices to enable communication between users and applications. Just like output devices, when developing augmented reality human-computer interaction, there are many different types of input devices to choose from, and some devices are more suitable for specific tasks than others. Augmented reality three-dimensional interactive input devices can be divided into discrete input devices, continuous input devices, and direct human input devices.

Discrete input device. This type of device generates an event at a time based on user actions, generating a simple value (boolean or element in a collection), often used to change the mode or start an action for discrete command interfaces such as keyboards, pinchglove, etc.

Continuous input device. Such devices track the user's continuous motion using different sensors including force, heat, light, electricity, and sound. The tracking point data includes information such as position, direction, or acceleration. Typical equipment such as three-dimensional mouse, electromagnetic tracker, force feedback gloves, data gloves, built-in sensor handle Wii Remote, camera, depth camera Kinect, Leap Motion, 3D camera RealSense and so on.

Voice and physiological signal sensing devices. Such devices mainly include voice input, bioelectric input and brain wave input. Representative voice input products include smart speaker Google Home and voice assistant Amazon Echo. The bioelectric input device mainly interacts by reading a human muscle or nerve signal change by a bioelectric sensor, and is mainly used in a smart wearable device. The brainwave input device monitors the activity of brain waves by EEG signals such as EEG signals. Representative products include NeuroSky MindWave mindset, EmotivInsight capable of manipulating drones, BrainLink with ability to concentrate, and EmotivEpoc which can imitate expressions. These direct human input devices complement other input channels and are an integral part of the ideal augmented reality interaction.

In augmented reality human-computer interaction design, the selection of output/input devices needs to consider the requirements of specific interaction technologies, the mutual constraints between input devices and output devices, and the complementarity between multi-channel interactions. In practice, the cost is often the biggest factor, and it is also necessary to consider whether the design of the interaction technology is limited by the given device, whether it is necessary to purchase advanced equipment for implementing the interaction technology, and whether it is necessary to create a new interaction device for the interaction technology.

4. Human-computer Interaction Technology in Augmented Reality

4.1 Three-dimensional Interactive Technology

Three-dimensional interaction is the most important way to interact with augmented reality. Compared with two-dimensional interaction, Augmented Reality 3D interaction provides more freedom of operation, interactive tasks are more complex, and interactive interface design has more design space, so new interaction metaphors and techniques are needed.

Metaphor can implicitly indicate the relationship between subject and metaphor, and is the way to construct the internal relationship between the two. The metaphor in the interface design is to use the similarity of the real things familiar to the user, to trigger people's emotional resonance, thereby reducing the complexity of the human-computer interaction interface. In three-dimensional interaction, the interaction metaphor maps the spatial direction/position information of the input device or the discrete button state to the operation of the virtual space to complete a specific interaction task. Different interaction metaphors provide different ways of interacting. According to the different metaphors of the input system, the existing three-dimensional interaction modes can be divided into direct mapping and indirect mapping.

Direct mapping interaction refers to directly mapping the location/spatial information input by the device to the operation action of the hand or device in the virtual space. The main metaphorical methods include ray projection metaphor, virtual hand metaphor, and so on.

Indirect mapping interaction refers to mapping the information input by the device into gestures, controlling the proportion of the scene space through gestures, and then completing the interactive tasks in the new proportional space. The main metaphorical methods include image plane metaphors and puppet metaphors.

Three-dimensional interactive technology has been applied in some areas such as video games, interactive experiences and so on. Generally speaking, the current three-dimensional interaction methods mainly have such problems as the spatial scope of the three-dimensional operation, the difficulty of controlling the extension length of the virtual hand, the effective control of the degree of freedom, and the presentation of the proportional space.

4.2 Gesture Interaction Technology

Gesture interaction refers to the posture or movement of a human hand carrying certain information, generally divided into static and dynamic gestures. A static gesture is usually a momentary gesture of the hand, equivalent to a point in a discrete space. A dynamic gesture, that is, a series of consecutive actions of a hand, can also be thought of as a sequence of multiple static gestures in a sequence of time and space, equivalent to multiple points or trajectories in a contiguous space.

According to whether it has purpose, the hand movement can be divided into two categories: gestures and unconscious movements. Gestures are purposeful and can be divided into operational gestures and communicative gestures. Manipulating gestures is the natural behavior of grabbing or discarding objects. They are not created specifically for interaction. Communication gestures are suitable for human-computer interaction, and can convey a special meaning in gestures, such as traffic police to direct traffic through their special gestures. From the current research, human-computer interaction tends to be more natural, and interactive methods such as natural posture have important theoretical and practical value. Traditional gesture recognition methods can be divided into three categories according to the principle: based on computer vision; based on data gloves; based on somatosensory devices.

Gesture interaction and augmented reality technology are hot research fields in recent years, and a large number of augmented reality applications require the support of gesture interaction technology. The use of the recognized three-dimensional coordinates of the fingertips can express various complicated gesture operations, achieving more natural and humanized human-computer interaction.

4.3 Handheld Mobile Device Interaction Technology

With the advancement of electronic information technology, the computing power and graphics display capability of handheld mobile devices represented by smart phones have been continuously enhanced, and various types of sensors such as cameras, GPS, accelerometers, gyroscopes, and electronic compasses have been integrated. Improves the device's ability to sense the environment and user behavior. In addition to focusing on the user's interaction with these smart mobile devices themselves, people have also noticed the three-dimensional interaction capabilities of smart handheld mobile devices, which are used in the field of enhancement/virtual reality [5] and pervasive computing. The system based on the interaction of handheld devices has its own characteristics, but a common idea is to get rid of the traditional desktop interaction and realize the interaction between the user and the environment. The core of the system is the design problem of three-dimensional interaction based on handheld mobile devices. There are two main types of interactive metaphors based on handheld mobile devices: information lens metaphors and direct pointing controller metaphors.

Information lens metaphor is currently the most used method, especially in the field of augmented reality, the core idea is to calculate the relative position of the handheld device relative to the observed scene (marked digital information) in real time (by position sensor or visual method), The labeled digital information is superimposed on the image captured by the camera in real time, and subsequent operations will operate in 2D based on buttons, pens or touch screens on the handheld device. The information lens metaphor originated from the development of the Chameleon system by Fitzmaurice[6], which proposed the concept of a space-aware display. That is, using a handheld device with a position tracker to sense digital information superimposed on the space.

The direct pointing metaphor is to capture the position of the handheld mobile device relative to the spatial display, thereby manipulating the spatial display through the handheld device. In recent years, with the advancement of display technology, large-scale display devices for the public are used in various occasions, but there is a certain problem in interacting with them: the traditional mode of using the mouse and keyboard requires the user to be in a fixed position, which limits the problem. The screen acts as a utilization of a common interaction space. By utilizing the mobility of the handheld mobile device, the user can operate the remote screen at a fixed position, which is convenient for multiple users and mobile use. Will greatly improve the efficiency of the use of large screens.

4.4 Eye Movement Interaction Technology

Eye movement measurement technology has evolved over a hundred years, from the original visual observation method to the later mechanical recording method, to a series of new techniques for accurate measurement. At present, mainstream eye movement measurement methods are mainly divided into two categories, one based on hardware and the other based on software[7-8].

Hardware-based measurement method. One of the classic methods is to fix the head orientation by means of a helmet or head fixing bracket. Using an eye camera, the infrared light reflected from the cornea and pupil of the human eye is recorded, and the infrared light is emitted by the infrared light source. Sight line tracking is achieved by analyzing changes in line of sight by analyzing changes in infrared light. Whether it is forced, wearable or not, is the difference between the line-of-sight tracking instruments. The accuracy of each instrument ranges from 0.1° to 1°.

Software-based measurement method. This method uses image processing technology to locate and track the face or eye image, extract the pupil position, and estimate the position of the user's line of sight. With the improvement of computer vision software and hardware technology, and its non-interference characteristics for users, this natural measurement method is gradually applied to the actual.

Eye movement is an optional or complementary human-computer interaction input channel. Compared with other input methods, it has the following advantages: increasing the input bandwidth of human-computer interaction; improving the naturalness of interaction; improving the input efficiency of human-computer interaction; Personalized needs; eliminate the fatigue caused by traditional human-computer interaction.

The main problem that eye movements need to overcome is the Midas Touch problem. Midas is a king in Greek mythology, and whatever he touches, that thing will turn into gold. In eye-based human-computer interaction, this problem refers to the eye movement in the active way of the target selection operation, the system can not distinguish whether the user is deliberately selecting an object, or just look at whatever, the user's random line of sight may cause an error. action. Therefore, it is very necessary for the system to distinguish between the user's intentional choice and unintentional choice.

4.5 Voice Interaction Technology

The process of voice interaction includes speech recognition, natural language understanding, dialogue management, natural language generation, and speech synthesis. The dialogue management includes dialogue state tracking and dialogue scheme selection, which is equivalent to the decision-making layer. Natural language understanding and natural language generation constitute the natural language processing, which is equivalent to the cognitive layer. Speech recognition and speech synthesis are similar to the function of the perception layer.

Voice input is a natural input method that combines different types of input techniques to create a more coherent and natural interface. Voice input becomes a valuable tool in the augmented reality user interface if the functionality is appropriate, especially if both of the user's hands are occupied. Voice has many desirable features: it frees the user's hand; uses an unused input channel; allows efficient and accurate input of large amounts of text; is a completely natural and familiar way. In the augmented reality user interface, voice input is especially suitable for non-graphical command interaction and system control, that is, the user requests the system to perform a specific function, change the interaction mode or system state by issuing a voice command.

4.6 Force/tactile Interaction Technology

When the user completes some touch, hit, press, etc. in the virtual environment, the virtual object cannot be touched, and thus the immersion is lacking. The force tactile feedback device is a device that provides tactile and force sense information to the experiencer and has two-way information transmission capability[9], which can greatly improve the realism and immersion of the experiencer in human-computer interaction.

4.6.1 Active Force/tactile Reproduction Device

Currently, most force tactile reproduction devices are active: exoskeleton and stationary devices; data gloves and wearable devices; point interaction devices and specialized devices. Although these force/tactile rendition devices can provide a large range of force tactile feedback to the operator, they are generally bulky because the force generation and control are mainly based on the output control of active actuators such as electric, pneumatic, hydraulic and the like. Problems such as heavy weight and high friction make it easy for the operator to feel the illusion when he feels the virtual object. In addition, it is difficult for the motor to work in a locked state with a large output, and the feeling of hard contact cannot be generated. However, the pneumatic and hydraulic actuators have serious nonlinear interference, and it is difficult to accurately control the magnitude of the force feedback. These problems hinder the widespread use of force/tactile rendering techniques in augmented reality.

4.6.2 Passive Force/tactile Reproduction Device

The passive force/tactile feedback system is energy dissipative, so it has stability characteristics, does not cause damage to the operator, and is much larger than the range of the same volume of the main power / haptic feedback system feedback, so it is easy to force The tactile reproduction device is made small and light, overcoming the disadvantages of the active device. Electro-/magnetic rheological fluids are liquid smart materials that can be obtained smoothly and quickly by simply selecting the applied voltage/flow, any state between the liquid and solid state of the electro-/magnetic rheological liquid[10]. In recent years, people have begun to study the dynamic/tactile reproduction technology based on electrorheological and magnetorheological changes.

4.7 Multi-channel Interaction Technology in Augmented Reality

Multi-channel interaction means that users can sense product information and interact with them through multiple different channels or different forms of the same channel, and pursue human-machine dialogue in a natural, efficient, parallel, and collaborative manner. In the multi-channel user interface, users can work with computer systems using natural interaction methods such as voice, gestures, eyes, expressions, lips, etc. People and machines are regarded as active participants in information exchange, input channels. There are many ways of serial/parallel, complementary/independent, and human-computer interaction moves closer to the form of human-to-human interaction, which greatly improves the naturalness and efficiency of interaction[11], which will be the future augmented reality man-machine. The mainstream form of interaction.

Using multi-channel interaction in augmented reality has the following advantages: Reduce the coupling; Reduce errors and fixes.; Flexibility and complementary behavior; Reduce cognitive load.

5. Conclusion

Thanks to the development of graphics technology and display technology, augmented reality technology has made great progress, especially the development of helmet technology and the combination of augmented reality and the Internet and artificial intelligence, making augmented reality become a new type of personal computing platform. Possibly, this has prompted people to focus on more efficient and adaptable interaction technologies to address the diverse needs of augmented reality applications. From the current state of the art, human-computer interaction in the augmented reality environment also faces the following challenges: natural interaction behavior and state perception in the augmented reality environment; augmented reality environment understanding of human interaction behavior and intention; augmented reality interaction The adaptive problem of output feedback. In the future, the interaction between machines and people can be like the daily interaction between people, so that users have less or even no special training and memory, and they can use the machine with their existing experience.

References

[1] Zhao Yonghui. Survey of Human-Computer Interaction Research[J]. China Computer& Communication, 2017, (23): 24-25, 28.

[2] Mark green,Robert jacob. SIGGRAPH '90 Workshop report: software architectures and metaphors for non-WIMP user interfaces[J]. ACM SIGGRAPH Computer Graphics, 1991, 25(3): 229-235.

[3] Jacob R,Girouard A,Hirshfield L, et al. Reality-based Interaction: a Framework for Post-wimp Interfaces[C]//Proc. Chi 2008, [S.l.]: [s.n.], 2008: 201-210.

[4] Wang Dangxiao, Jiao Jian, et al. Computer Haptics: Haptic Modeling and Rendering in Virtual Reality Environments[J]. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(6): 881-895.

[5] Feng zhou,Duh,H.b.-l., et al. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR[C]//Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on, 2008: 193-202.

[6] George w. fitzmaurice. Situated information spaces and spatially aware palmtop computers[J]. Communications of the ACM, 1993, 36(7): 39-49.

[7] Sheena D. Survey of Eye Movement Recording Methods[J]. Behavior Research Methods & Instrumentation, 1975, 7(5): 397-429.

[8] J. k. jacob R,Karn K. Eye Tracking in Human-computer Interaction and Usability Research: Ready to Deliver the Promises[Z]. [S.l.]: [s.n.], 2003: 573-605.

[9] Lu Xiong， Chen Xiaoli, et al. Haptic rendering methods for natural humancomputer interaction: A review[J]. Chinese Journal of Scientific Instrument, 2017, 38(10): 2391-2399.

[10] Dai Jinqiao, Wang Aimin. Application of the SmartMaterials in the Haptics[J]. Measurement & Control Technology, 2007, 26(7): 3-6.
[11] Jiang Min， Han ChaoYan. Design & Research on Multimodal Human-machine Interface[J]. Development ＆ Innovation of Machinery ＆ Electrical Products, 2009, 22(3): 24-25.