



Research on the Development of Natural Human-Computer Interaction for Mobile Terminals

Qing Zhang¹ and Xiaoyong Lin²(✉)

¹ Computer Engineering College, Jimei University, XiaMen, China

² Xiamen University of Technology, XiaMen, China

qingzh_xm@163.com

Abstract. Human-Computer Interaction (HCI) technology, as an important part of computer systems, has developed rapidly in computer science. It has experienced a transition from humans adapting to computers, to computers constantly adapting to humans. With the development of human-computer interaction, users are more and more inclined to use natural communication methods such as natural language, gestures, and vision instead of traditional keyboard and mouse input. In today's era of the mobile Internet, mobile terminals have been widely used. Due to their portability and mobility, users expect their interactions with mobile terminals to be smooth and natural. The natural human-computer interaction of mobile terminals has become a research hotspot. Through the analysis of the evolution process of mobile terminal human-computer interaction, combined with the latest human-computer interaction technology, this paper discusses the hot issues in mobile human-computer interaction and concludes that the future human-computer interaction of mobile terminals has four development directions: 1) natural human computer interaction, 2) augmented reality interaction, 3) multi-modal fusion, 4) affective computing.

Keywords: Natural human-computer interaction · Mobile Terminals · AR · Multi-modal fusion · Affective Computing

1 Introduction

The term “mobile terminals” refers to computer devices that can be used in the mobile, that can provide digital information services or exchange data and information through the network. Mobile terminals include mobile phones, laptops, tablets, and other smart terminal devices.

With the emergence of the 5G network, mobile communication is becoming more and more broadband. In addition, most smart mobile terminals have strong computing power and have rapidly transformed from the original mobile network terminal to the key entrance of Internet business, becoming the main innovation platform in the mobile Internet era.

China Internet Network Information Center (CNNIC) released its 45th statistical report on the development of China's Internet. It shows that by March 2020, the number of Internet users in China had reached 904 million, including 897 million mobile Internet users accounting for 99.3%. Mobile Internet service scenarios are constantly enriching, the scale of mobile terminals is increasing rapidly, and the amount of mobile data is continuously expanding. It shows that Internet access through mobile terminals has become the most important access to the Internet. Human computer interaction for mobile applications has also become a research hotspot.

From the evolution process of mobile phone human-computer interaction, mobile human-computer interaction experienced the following three stages:

1. The human-computer interaction interface in the form of characters represented by MOTOROLA.

In 1993, Motorola launched the Motorola 3200 in China, which only includes voice and text messaging functions. Since then, the mobile phones launched by various mobile phone manufacturers have used physical keyboard for input, including numbers 0–9 and several function keys. The mobile phone has simple functions, simple interaction mode and poor user experience.

2. Graphical user interface represented by Apple iPhone.

The advent of Apple's iPhone in 2007 is a revolutionary product across the ages. The traditional physical-digital keyboard disappeared and was replaced by the touch screen interface. Instructions and information can be input through the direct touch of fingers or pens with the screen. The new touch screen interaction mode greatly facilitates the user's rapid human-computer interaction. The smooth operation experience of the iPhone also makes designers pay more attention to user experience, and user centered human-computer interaction design has become the mainstream.

3. Multi-modal human-computer interaction interface.

With the development of multi-modal human-computer interaction technology, visual, auditory, tactile, and other sensory channels are used in human-computer interaction such as gesture interaction, voice interaction, and expression interaction. Users can interact with mobile phones in a variety of ways, which greatly improves the user experience.

At present, promoted by the new generation of information technology clusters such as artificial intelligence, big data, cloud computing, VR/AR, etc., the natural human-computer interaction technology of mobile terminals has also entered a new development stage.

In the second part, the directions of human-computer interaction in mobile situations will be discussed from four research hot spots.

2 Research Review

2.1 Natural HCI

At present, the human-computer interface of mobile terminals mainly adopts graphical user interface (GUI). The traditional graphical user interface uses keyboard or mouse to

input user instructions. Users need to learn the operation methods set by software developers, and complete the interaction process according to the preset operation process in the operation process, which make users cost time for learning and is an unnatural human-computer interaction.

Natural human-computer interaction refers to a process in which users only use the existing cognitive habits and familiar behavior patterns to interact with the computer when they interact with the computer. It is usually imprecise and a natural behavior.

The goal of natural human-computer interaction is to get rid of the shackles of mouse and keyboard, allow users to use their own senses and existing life experience to operate, and reduce the cost and burden of learning as much as possible.

The common natural interaction technologies include multi-touch, gesture recognition, expression recognition, voice interaction and eye tracking. With the development of natural interaction technology, one or more natural interaction technologies are used in the interaction of mobile terminals, forming the prototype of natural user interface.

However, there are some usability problems in the use of natural user interface, such as limited use scenarios, lack of functional visibility, cognitive differences and so on. These usability problems will increase the learning cost of users and turn natural interaction into unnatural. Obviously, it is not enough to only apply natural interaction technology to human-computer interaction. From another aspect, interaction design should also adopt user centered design (UCD) method, design a natural interaction mode with the best user experience by considering user psychology, user habits, user types and use scenarios.

Besides, with the continuous breakthrough of brain-computer interface technology [5], it will be possible to directly control the computer with ideas, which will completely change the form of human-computer interaction in the future. Natural human-computer interaction will develop from tangible interface to invisible interface. The best interaction is natural, and the best interface is no interface [14]. The human-computer interaction of mobile terminals will gradually move towards a more humanized, more intelligent and more user-friendly level of natural experience.

2.2 HCI with AR Scenarios

Augmented Reality (AR) is developed on the basis of Virtual Reality (VR), it emphasizes the combination of virtual and real. It integrates the real environment of the real world and the virtual environment generated by computer in real time, to bring users a relatively realistic comprehensive feeling in hearing, vision and touch, and realize the natural integration of human and environment. Augmented reality oriented human-computer interaction has the characteristics of virtual reality superposition, three-dimensional, real-time interaction.

With the rapid development of mobile terminal devices, the products represented by high-performance smart phones and wearable devices (smart glasses, etc.) provide a carrier for the practical application of augmented reality in mobile terminals. If the AR system can be integrated into a mobile phone, the camera is responsible for collecting images, and the processing unit is responsible for analyzing and reconstructing the images, to realize the alignment of the coordinate system and the fusion calculation of

the virtual scene. The processed images will be displayed on the screen of the mobile phone, to realize the realistic enhancement effect.

AR provides users with a new interaction mode, including motion capture, tactile feedback, eye tracking, EMG simulation, etc. Among them, eye control interaction technology could introduced into mobile augmented reality. Users only use line of sight to operate instead of VR device, interact with mobile phone by eyes, that is especially suitable for human-computer interaction of mobile terminal AR [2, 3].

An AR device that could replace the mobile may be developed in the future, make all products with screens a thing of the past. It could allow consumers to scroll through applications without obscuring visibility of the real world. Dreaming even bigger, the highest level of augmented reality is "what you see is what you think, what you think is what you can", to achieve a more immersive interactive experience.

2.3 HCI with Multi-modal Fusion

Mobile terminals are portable, users expect to input and output instructions quickly and conveniently, which makes it more urgent for people to improve the efficiency of human-computer information exchange through visual, auditory, tactile and other interactive ways. At present, although mobile terminals adopt various human-computer interaction modes of sensing modal technology, the information recognition and processing of each modal are mostly separated, and the real intention of users may not be accurately obtained.

Intelligent human-computer interaction requires the fusion of multiple sensory information, namely multi-modal fusion. When you see a picture, text could be generated, when you see the text, pictures and videos could be imaged. An agent should be able to complete the modal transformation between vision and semantics. Multi-modal human-computer interaction is actually the simulation of natural interaction between people. It transplants the interaction mode between people into the interaction between human and computer, aiming at reducing the gap between human and computer, and creating a natural and harmonious human-computer environment.

At present, multi-modal user interface uses new technologies such as eye tracking, speech recognition, gesture input, etc., users can use multiple sensory modals to conduct human-computer interaction in a natural, parallel and cooperative method. By fusing multi-modal accurate and imprecise information, the system can quickly capture the user's intention. This relies on the technology of "multi-modal deep learning", which enables the agent itself to understand the multi-modal signal. It needs to accommodate the auditory, visual and sensing signals for unified thinking, so that the machine can carry out multi-modal collaborative learning and truly "smart". The mobile terminals based on artificial intelligence technology will be a robot with various sensors such as vision, hearing, smell and taste [11–13].

In addition to perceptual intelligence, mobile applications on mobile phones can also connect to the cloud brain through the network and have cognitive intelligence. At present, most of the solutions of mobile artificial intelligence rely on cloud computing, but in the application scenarios which need high real-time response, the computing of mobile terminals is also essential. In addition, security and privacy also need to use the advantages of terminal computing to achieve. With the upgrading of mobile artificial

intelligence chip and optimization of algorithm, part of the computing and processing functions of AI should be migrated from cloud to mobile terminals.

2.4 HCI with Affective Computing

Even if computer is given a variety of intelligents, it is still unable to understand and adapt to human emotions, it unable recognition human emotion, also unable to express emotions.

In 1997, Professor Rosalind W. Picard who is the founder of the Affective Computing Research Group at the MIT Media Lab proposed the concept of affective computing, which includes emotion recognition, emotion representation, Emotion Modeling and emotional interaction [15]. Using emotional computing, it is expected that computers will have the ability to observe, understand and generate various emotions similar to human beings, and finally make human-computer interaction as natural as human interaction.

Affective computing is a highly integrated research field, which combines computational science, psychological science and cognitive science. It can be used in the process of human-computer interaction. By studying the emotional characteristics of human interaction and human-computer interaction, a human-computer interaction environment with emotional feedback is constructed. The human-computer interaction not only has high perception and cognitive intelligence, but also has high emotional intelligence. In the future the computer will have high EQ, which can effectively solve the situation perception, emotional understanding and emotional expression in human-computer interaction, and make reasonable response.

At present, affective computing is still in its infancy, and most of the research hotspots and achievements are reflected in the level of emotion recognition. For example, in facial expression recognition or natural language processing, machine learning and convolution neural network can be used to identify anger, disgust, fear, happiness, calm, sadness, surprise and so on. Most of the intelligent applications based on mobile terminals use the powerful computing power of the cloud to realize emotion recognition. This type of emotion recognition needs to be from the mobile terminals to the cloud, and then back to the mobile terminals from the cloud. Under the limitation of network bandwidth, there may be delay. However, with the advent of 5G era, the cloud based emotion recognition which highly real-time response is still worth looking forward to.

On the other hand, during the continuous evolution of intelligent mobile terminals, it is also trying to improve the traditional processor computing architecture so as to support the high-speed computing and low-power consumption required by machine learning. Emotion recognition, emotional interaction will become the instinct of mobile terminals.

Emotional information is often expressed from multiple dimensions such as language, voice intonation, facial and limbs in face-to-face communication. In the process of human-computer interaction, emotional feature recognition also needs to be calculated from multiple dimensions, such as text emotion analysis, facial expression recognition, speech emotion recognition, posture recognition, and even through physiological pattern recognition, such as skin electric response, breathing, heart rate, body temperature, brain wave and so on [10]. Multi-modal emotion information fusion and combining the context information of the situation at that time would make computer recognize and understand the human emotion.

Human-computer interaction with affective computing is shown in the Fig. 1 below.

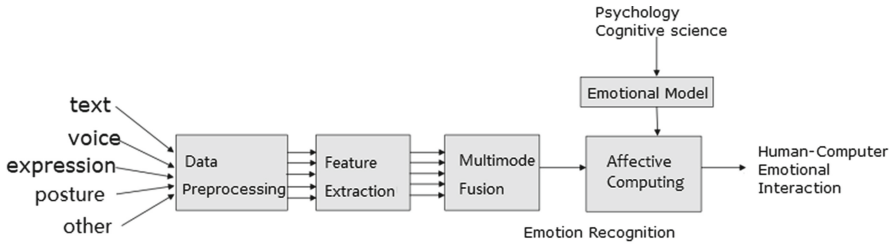


Fig. 1. HCI with affective computing

In addition, how to establish an emotional model based on psychology and cognitive science to express the relationship between emotion, cognition and will, which is suitable for machine implementation, will also be a great challenge.

3 Conclusion

The pervasiveness of mobile terminal services and applications fundamentally changed the way we access information and communicate with each other. With the continuous breakthrough of human-computer interaction technology, mobile human-computer interaction presents a trend of continuous development towards natural human-computer interaction, VR/AR, multi-modal fusion, and emotional human-computer interaction. The human-computer interaction of mobile terminals is gradually moving towards a more human-oriented, more intelligent, and more natural experience.

We can imagine the future through the comprehensive calculation of the user emotional model and generate empathy and emotionally interactive content through the addition of emotional factors to enhance user experience, establishing human-computer interaction with empathy, situational awareness, and natural harmony.

Acknowledgements. Major Research Topics of Social Science Base in Fujian Province: “Research on the future media industry and the development of strategic emerging industries in Fujian Province” (fj2018jdz055), 2018–2020.

References

1. Yi, X., Yu, C., Shi, Y.C.: Bayesian method for intent prediction in pervasive computing environments (in Chinese). *Sci. Sin. Inf.* **48**, 419–432 (2016). <https://doi.org/10.1360/N112017-00228>
2. Li, F.Y., Feng, J.P., Fu, M.S.: Research on natural human-computer interaction in virtual roaming. *J. Phys. Conf. Ser.* **1518**(1), 012022 (2020). <https://doi.org/10.1088/1742-6596/1518/1/012022>

3. Su, G.E., Sunar, M.S., Ismail, A.W.: Device-based manipulation technique with separated control structures for 3D object translation and rotation in handheld mobile AR. *Int. J. Hum. Comput. Stud.* **141**, 102433 (2020). <https://doi.org/10.1016/j.ijhcs.2020.102433>
4. Li, X., Zhang, M.: Emotion analysis for the upcoming response in open-domain human-computer conversation. In: U, L.H., Xie, H. (eds.) *APWeb-WAIM 2018*. LNCS, vol. 11268, pp. 352–367. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01298-4_29
5. Robert, L., Daniel, P.M.: Brain-computer interfaces and virtual reality for neurorehabilitation. In: *Handbook of Clinical Neurology*, vol. 168 (2020). <https://doi.org/10.1016/B978-0-444-63934-9.00014-7>.
6. Zhang, X.Y., Ban, X.J., Cheng, Z., Liu, T.: Modeling and recognition of human limbs cooperative interaction based on Random Increased Hybrid Learning Machine. *Procedia Comput. Sci.* **147**, 198–202 (2019). <https://doi.org/10.1016/j.procs.2019.01.222>
7. Alfaro, L., Linares, R., Herrera, J.: Scientific articles exploration system model based in immersive virtual reality and natural language processing techniques. *Int. J. Adv. Comput. Sci. Appl.* **9**, 254–263 (2018). <https://doi.org/10.14569/IJACSA.2018.090736>
8. Bachmann, D., Weichert, F., Rinkenauer, G.: Review of three-dimensional human-computer interaction with focus on the leap motion controller. *Sensors* **18**(7), 2194 (2018). <https://doi.org/10.3390/s18072194>
9. Le, H.Y.: Modeling human behavior during touchscreen interaction in mobile situations. In: *MobileHCI 20'16 Adjunct*. ACM (2016). <https://doi.org/10.1145/2957265.2963113>, 978-1-4503-4413-5/16/09
10. Patanè, A., Kwiatkowska, M.: Calibrating the classifier: siamese neural network architecture for end-to-end arousal recognition from ECG. In: Nicosia, G., Pardalos, P., Giuffrida, G., Umerton, R., Sciacca, V. (eds.) *Machine Learning, Optimization, and Data Science. LOD 2018. Lecture Notes in Computer Science*, vol. 11331, pp. 1–13. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-13709-0_1
11. Cuzzocrea, A., Mumolo, E., Grasso, G.M.: An effective and efficient genetic-fuzzy algorithm for supporting advanced human-machine interfaces in big data settings. *Algorithms* **13**(1), 13 (2019). <https://doi.org/10.3390/a13010013>
12. Tsiourti, C., Weiss, A., Wac, K., Vincze, M.: Multimodal Integration of emotional signals from voice, body, and context: effects of (in)congruence on emotion recognition and attitudes towards robots. *Int. J. Soc. Rob.* **11**(4), 555–573 (2019). <https://doi.org/10.1007/s12369-019-00524-z>
13. Hobeom, H., Won, Y.S.: Gyroscope-based continuous human hand gesture recognition for multi-modal wearable input device for human machine interaction. *Sensors (Basel, Switzerland)* **19**(11) (2019). <https://doi.org/10.3390/s19112562>
14. Krishna, G.: *The Best Interface is No Interface: The Simple Path to Brilliant Technology*. Pearson Education Inc., New York (2015)
15. Rosalind, W.: *Picard: Affective Computing*. The MIT Press, Cambridge (1997)
16. Liu, G., Wang, Y., Orgun, M.A.: Finding K optimal social trust paths for the selection of trustworthy service providers in complex social networks, *IEEE Trans. Serv. Comput.* **6**(2) (2013)
17. Liu, G., Wang, Y., Orgun, M.A.: Optimal social trust path selection in complex social networks. In: *Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2010)*, pp. 1391–1398 (2010)
18. Liu, G., et al.: MCS-GPM: multi-constrained simulation based graph pattern matching in contextual social graphs. *IEEE Trans. Knowl. Data Eng.* **30**(6), 1050–1064 (2018)
19. Liu, G., et al.: Multi-constrained graph pattern matching in large-scale contextual social graphs. In: *IEEE 31st International Conference on Data Engineering (ICDE 2015)*, pp. 351–362 (2015)