Human Factors and Systems Interaction, Vol. 154, 2024, 134–144
https://doi.org/10.54941/ahfe1005361

Design and Implementation of Gesture Interaction for a Command and Control System Based on Multimodal Data Fusion

Mengmeng Gao, Deqiang Fu, and Ning Li

Marine Human Factors Engineering Laboratory, China Institute of Marine Technology and Economy, 70 xueyuan south road, haidian district, Beijing, China

ABSTRACT

With the continuous improvement of people's demand for interactive experience, gesture interaction technology, as a natural and intuitive way of human-computer interaction, has attracted more and more attention and research. However, the research and application of gesture interaction in command and control system is still lacking. Firstly, the characteristics of command and control system and the key technologies of gesture interaction are investigated in this study. Then a series of operating gestures that meet the application scenarios of the command and control system are designed, and a gesture recognition algorithm based on the fusion of image data and skeletal data is proposed. By integrating with the command and control system, the function of controlling the command and control system through gestures is realized. Finally, the feasibility and effect of gesture interaction technology in the application scenarios of command and control system are evaluated through experiments. The experimental results show that the application effect of the technology in the command and control system is good, which can reach 94.89% under typical tasks, improving the interaction efficiency and interaction experience.

Keywords: Human-computer interaction, Command and control system, Gesture recognition, Multimodal human-computer interaction

INTRODUCTION

Traditional input devices, such as mice and keyboards, are becoming insufficient in meeting users' demands for more natural and intuitive ways of interaction. This gap has led to the emergence of gesture interaction technology, which utilizes human hand movements as input signals to facilitate interaction with computers and other electronic devices (Zhang, 2021). Gesture interaction technology stands out for its intuitive, natural, and convenient nature, offering users a more immersive and engaging experience. The recent advancements in artificial intelligence and hardware computing power have propelled the widespread adoption of this technology across various domains, including virtual reality (Yuan, 2022), gaming (Gupta, 2016), smart homes (Tan, 2020), and healthcare (Luo, 2020).

As a prominent application scenario, the command and control system holds immense significance in exploring the potential of gesture interaction technology. Traditional methods of operating command and control systems often involve complicated steps and a steep learning curve (Chen, 2019), posing challenges for efficient command execution. Therefore, this paper endeavors to explore how gesture interaction technology can revolutionize the operation of command and control systems, streamlining processes and ultimately enhancing command efficiency.

OVERVIEW OF THE COMMAND AND CONTROL SYSTEM

The command and control system is a human-machine interaction platform that integrates information synthesis, information display, and simulation. It can assist commanders in making decisions, issuing orders, and implementing commands. In today's complex battlefield environments, where information sources are abundant and situations evolve rapidly, traditional military command and control systems often rely on a singular, graphically-based interaction paradigm. However, this approach can impose significant cognitive loads on operators and hinder interaction efficiency. To meet the demands of the modern, high-intensity battlefield, the future of command and control systems must evolve towards a "human-centered" design philosophy that prioritizes natural and intuitive interaction methods. This is of great significance for improving the efficiency of command and control system.

RESEARCH ON GESTURE INTERACTION TECHNOLOGY

The gesture interaction system comprises four main components: the operator, gesture input device, gesture recognition algorithm, and the operated device or corresponding interface (see Figure 1).

Figure 1: Components of gesture interaction system.

Classification of Gestures

Gesture can be divided into static gestures and dynamic gestures based on the movement status of the hands (see Figure 2). Static gestures, also known as static hand shapes, are stationary hand postures without obvious movement or continuous actions, such as the thumbs-up gesture, fist gesture, V-shaped gesture, and OK gesture, etc. Dynamic gestures consist of a series of continuous action sequences, involving spatial displacement and shape changes of the hands over time, typically conveying specific meanings or information. Examples include waving gestures, zoom in/out gestures, sliding gestures, and tapping gestures, etc.

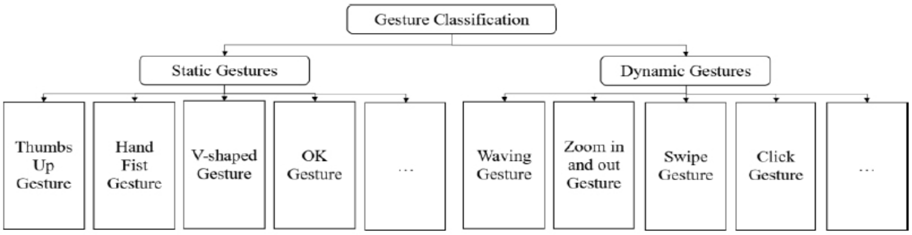

Figure 2: Classification of gestures.

Gesture Input Devices

Currently, the mainstream gesture input devices on the market can be divided into two major categories (see Table 1), namely optical capture devices and wearable capture devices (Liu, 2021).

Optical capture devices utilize sensors and cameras to capture user hand movements, typically employing technologies such as infrared and depth sensors. They offer the advantage of providing high precision and wide-ranging gesture tracking, thus acquiring rich three-dimensional information. However, they are susceptible to ambient light conditions, and poor lighting may affect capture effectiveness. Additionally, they are constrained by workspace limitations, requiring users to remain within the camera's field of view. Representative devices include Leap Motion developed by Ultraleap and Microsoft Kinect developed by Microsoft.

Wearable devices can be directly attached to users' body parts such as wrists, arms, or fingers, and they integrate sensors and data processing units to precisely capture gestures and movements from specific areas (Xie, 2015). Their advantages lie in portability and immediacy, allowing users to easily carry them without relying on bulky equipment and enabling instant capture of user dynamics thanks to their close proximity to the body, thereby significantly reducing data processing delays. Furthermore, wearable devices offer users enhanced flexibility, enabling them to operate freely without spatial constraints. However, these devices also have certain limitations, such as potentially lower accuracy in capturing subtle gesture variations and a more restricted range of application scenarios, which may prevent them from demonstrating the versatility and flexibility of optical devices in diverse environments. Nevertheless, in specific situations, wearable devices still exhibit their practical value. Representative devices include the Myo Armband gesture control armband developed by Thalmic Labs in Canada and the NImble VR developed by CrunchBase.

Table 1. Comparison between optical capture devices and wearable devices.

Gesture input device	Advantages	Disadvantages	Representative equipment
Optical capture device	High precision Large-scope 3D information	sensitive to light Limited work space	Leap Motion Microsoft Kinect
wearable capture device	Portable and lightweight Live capture Flexibility	Flexibility Relatively low accuracy Limited work scenarios	Myo Armband NImble VR

Gesture Recognition Algorithms

With the continuous advancement of computer vision and deep learning technology, vision-based gesture recognition algorithms have rapidly emerged as the preferred method for gesture detection due to their significant advantages such as low cost and high efficiency. These algorithms mainly receive two types of input data: image data and depth data. In addition, numerous gesture detection algorithms based on multimodal data fusion have been derived.

Image data can be divided into static images and continuous image sequences. Static image detection involves gesture recognition using a single image, which is suitable for recognizing and classifying static gestures. An algorithm proposed by Liu Shuping et al. utilizes Histogram of Oriented Gradients (HOG) feature extraction for static gesture recognition. This algorithm first segments the hand based on a skin color model, then extracts hand HOG features, and finally inputs them into a support vector machine (SVM) classifier for gesture recognition. Traditional static gesture detection algorithms require manually designed feature operators, and the detection performance is highly dependent on the design of these features, leading to poor robustness and scalability. Currently, the mainstream detection method involves end-to-end detection algorithms based on deep learning. Yuan Rongshang (Yuan, 2019) proposed a gesture detection algorithm based on AlexNet, significantly improving detection accuracy. Continuous image sequence detection involves recognition using a series of consecutive images, which is suitable for identifying dynamic gestures or hand movements. This method captures dynamic features of gestures by analyzing changes in the image sequence. Nie Gang (Nie, 2019) proposed a three-frame difference method based on Gaussian background subtraction, achieving dynamic gesture recognition for input image sequences.

Wu Yuxia (Wu, 2014), for instance, extracted features from depth images and performed structural matching to recognize gestures, ultimately achieving gesture-controlled slide presentations. Skeleton data, on the other hand, represents the three-dimensional coordinates of hand joints obtained using depth sensors. This input form provides information about the positions and postures of hand joints and bones, aiding in describing the movements and actions of gestures more precisely. Zhao Feifei (Zhao, 2016) modeled gesture recognition based on skeleton data and integrated it with a browser to enable gesture-controlled web browsing functionalities.

In order to obtain more comprehensive gesture information, gesture detection algorithms have derived increasingly multimodal data inputs, such as the fusion of image and depth data, which combines RGB image data with depth image data for gesture recognition based on both color and depth information. Chang Zhilei (Chang, 2023) studied the use of RGB images to locate hand parts and then fused depth images to estimate and calculate key points on the hand, ultimately obtaining gesture recognition results. You Fengtao (You, 2022) obtained more accurate spatial information by fusing two-dimensional images with skeletal information, improving detection accuracy. Alternatively, infrared images can be fused with skeletal data to obtain more comprehensive gesture information.

In general, there are several shortcomings in gesture interaction for command and control systems at present: Firstly, there is a lack of designed gestures suitable for typical tasks in command and control systems. Secondly, there is insufficient research on algorithms capable of simultaneously detecting static and dynamic gestures, and the accuracy of detection algorithms for dynamic gestures is inadequate. Thirdly, there is a lack of validation for gesture applications in command and control systems.

DESIGN AND IMPLEMENTATION OF GESTURE INTERACTION TECHNOLOGY BASED ON A COMMAND AND CONTROL SYSTEM

Gesture Design

Due to the limited variety of existing gesture command libraries and their applicability only to specific scenarios, this study combines gestures with the command and control system applications to design and optimize a set of gesture libraries that better align with the usage habits of operators in command and control systems. This includes a range of common gestures such as zoom in, zoom out, move, and swipe. The schematic diagram of the gestures designed in this paper is shown in Figure 3. The first row depicts zoom in, continuous zoom in, zoom out, and continuous zoom out gestures. The second row shows left move, right move, up move, and down move gestures. The third row represents left swipe, right swipe, up swipe, and down swipe gestures. The aim is to provide a better user experience and higher interaction efficiency, meeting the needs of user groups for gesture interaction in the command and control scenarios.

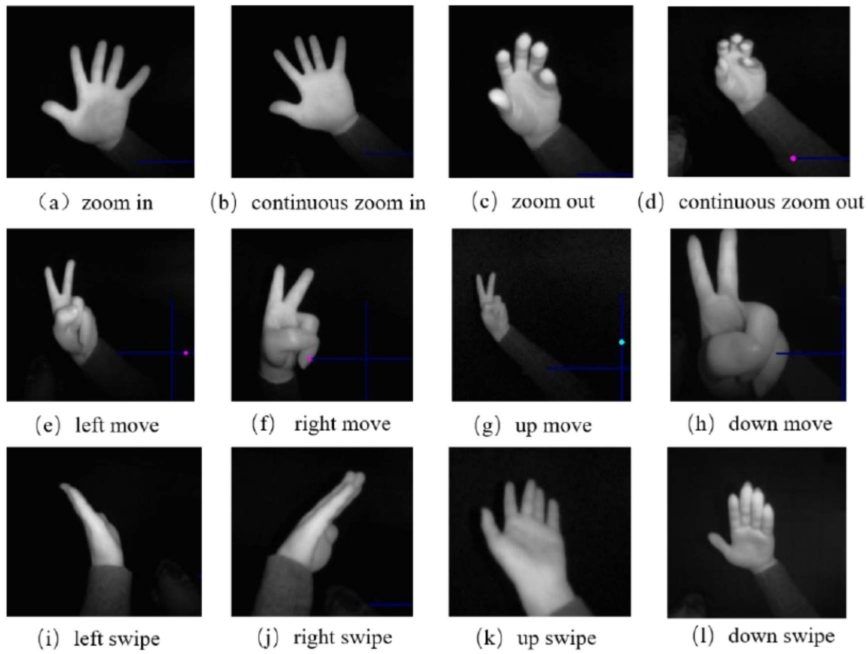

Figure 3: Schematic diagram of gesture command library.

Gesture Interaction Implementation Based on the Command and Control System

The architecture of the gesture interaction system in this study comprises a gesture recognition module, a system interface module, and a control module. The control module serves as a bridge between the gesture recognition module and the system interface module.

The gesture recognition module consists of a gesture data acquisition sensor and a gesture recognition model. To select appropriate hardware for gesture data collection, considerations such as device resolution, frame rate, and sensor type are necessary. Ultimately, Leap Motion was chosen as the collection device. Subsequently, this study constructed a gesture recognition model based on multimodal data fusion. Firstly, a static image-based hand shape recognition model was built using MobileNet. Secondly, a dynamic gesture recognition model based on skeletal sequences was constructed using a graph convolutional network. To improve recognition accuracy, this study fused image information and skeletal data to provide more information for the final decision, with the network structure depicted in Figure 4.

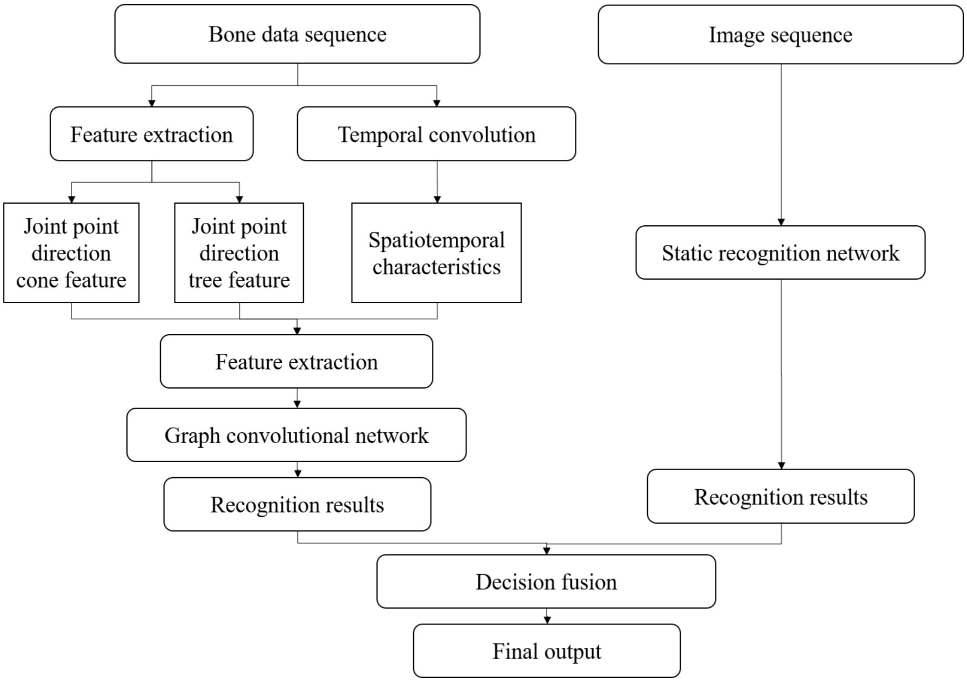

Figure 4: Gesture recognition model based on multimodal data fusion.

The system interface module is a set of background programs with a display interface developed based on the command and control background, primarily used to verify the feasibility and effectiveness of gesture interaction.

Considering the scalability of the system and to facilitate the integration of new functions in the gesture recognition module in the future, this study developed a control module to ensure smooth information transmission and collaborative work between the gesture recognition module and the system interface module. The control module includes an interaction interface between the gesture recognition module and the system interface module. It receives the gesture recognition results from the gesture recognition module, parses and sends them to the system interface module, and invokes the corresponding interface interaction response interface to complete the control operation of the gesture on the system interface.

Experimental Design and Implementation

To evaluate the feasibility and effectiveness of gesture interaction technology based on the command and control system, we conducted a gesture interaction experiment. Firstly, we recruited 15 participants and provided them with basic training on gesture operations. Next, we designed a series of experimental tasks that covered gesture interaction operations in different application scenarios. The overall testing process was repeated three times, and the test results were recorded each time.

In total, 10 types of gestures were tested, including "swipe up," "swipe down," "swipe left," "swipe right," "move down," "move up," "move right," "move left," "zoom in," and "zoom out." The testers were required to sit upright at the testing station and perform the gesture operations using their non-dominant hand (for right-handed people, the left hand was used for gestures, and for left-handed people, the right hand was used). When performing the gestures, they were instructed to keep their palm 5–10cm away from the table surface, within the field of view of the Leap Motion (see Figure 5).

Figure 5: Schematic diagram of gesture data collection.

For this experiment, a total of 10 test cases for gesture recognition accuracy were designed based on 10 types of gestures. When corresponding gestures were made on different display interfaces, there would be expected responses on the interface. Whether the gesture interaction was successfully completed was judged based on whether the system interface achieved the preset response. The specific gesture operations and expected interface responses are shown in Table 2.

Table 2. Gesture test cases.

Serial number	Gesture input	Instructions	Expected test results
1	"Wave up" gesture	Perform the "swipe up" gesture in the dual-screen interface	The pages displayed on the upper and lower screens are exchanged.
2	"Wave down" gesture	Perform the "swipe down" gesture in the dual-screen interface	The pages displayed on the upper and lower screens are exchanged.
3	"Left wave" gesture	Perform the "left swipe" gesture in an interface with multiple tab pages	Tab page switches tabs to the left.
4	"Swipe right" gesture	Perform the "right swipe" gesture in an interface with multiple tab pages	Tab page switches tabs to the right.
5	"Move down" gesture	Perform the "move down" gesture on a page with tables	The vertical scroll bar of the table slides down with the gesture.
6	"Move up" gesture	Perform the "move up" gesture on a page with tables	The vertical scroll bar of the table slides up with the gesture.
7	"Move right" gesture	Perform the "move right" gesture on a page with tables	The horizontal scroll bar of the table slides to the right with the gesture.
8	"Move left" gesture	Perform the "move left" gesture on a page with tables	The horizontal scroll bar of the table slides to the left with the gesture.
9	"Zoom in" gesture	Perform the "zoom in" gesture on a page with a map	The map follows the gesture process to zoom in.
10	"Zoom out" gesture	Perform the "zoom out" gesture on a page with a map	The map zooms out following the gesture process.

Analysis of Experimental Results

During the testing process, the 15 trained test users sequentially tested the 10 types of gestures. Each gesture was tested for 3 rounds, with the results recorded after each test. The total number of correct recognitions and the total number of tests were then tallied, with the ratio of correct recognitions to total tests serving as the recognition accuracy. The specific calculation formula for accuracy is as follows:

$$\text{Accuracy} = \text{TP/N}$$

Where TP represents the total number of correct recognitions according to the evaluation criteria in the test cases, and in this gesture recognition accuracy test, TP = 427; N is the total number of tests performed, and in this gesture recognition accuracy test, N = 15310 = 450. Therefore, the gesture recognition accuracy rate is 94.89%, and the detailed test situation is shown in Figure 6.

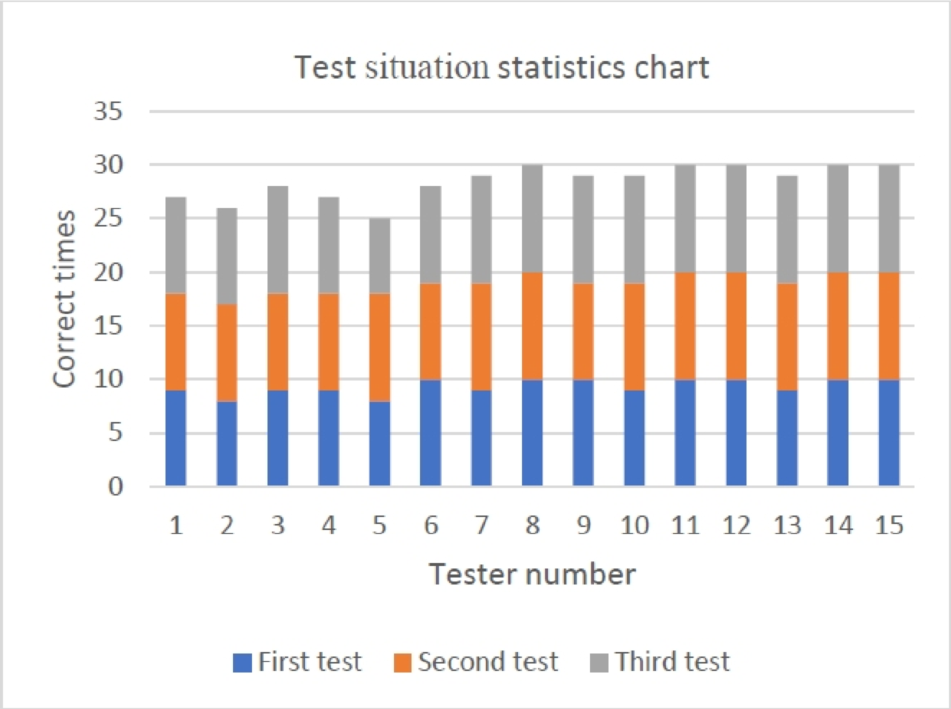

Figure 6: Statistical chart of gesture accuracy testing.

Through a series of tests, we comprehensively evaluated the accuracy performance of gesture interaction technology in various scenarios. Data analysis from the tests shows that gesture interaction technology based on command and control systems generally demonstrates satisfactory accuracy in most cases. Participants were able to quickly grasp gesture operation techniques during the testing process and highly praised the interactive experience provided by the technology, indicating a high level of satisfaction. However, we also observed that in certain specific contexts, such as executing complex gestures or encountering unfamiliar application scenarios, the accuracy of gesture interaction technology still needs improvement. Addressing these issues, we will continue to strive to optimize the technology to further enhance gesture recognition accuracy and user experience.

CONCLUSION

Addressing the shortcomings of gesture interaction in command and control systems, the main contributions of this study are as follows:

Firstly, we designed gesture actions suitable for typical tasks in command and control systems, addressing the lack of applicability in gesture control actions for command and control systems.

Secondly, we proposed a gesture detection model based on the fusion of skeleton data and infrared data, resolving the issue of low gesture recognition accuracy.

Thirdly, we selected 10 typical tasks in command and control systems to validate the model, evaluating the effectiveness of gesture interaction technology based on multimodal data fusion in command and control systems. Experimental results showed that the gesture recognition accuracy reached 94.89%, indicating the favorable application effects of this technology in command and control systems.

Furthermore, through surveys on participants' satisfaction with this interaction method, the majority perceived this interaction method as having great potential, being more intuitive, natural, and efficient than traditional human-computer interaction methods.

These research findings can support the development of command and control systems based on multimodal human-machine interaction.

ACKNOWLEDGMENT

Thanks for the technical support provided by the Equipment Advance Research Project Foundation of China (No. 50912020105).

REFERENCES

Chang, Z. (2023). Research and system design of dual-hand gesture recognition method based on RGB-D [Master's thesis, Inner Mongolia University]. https://doi.org/10.27224/d.cnki.gnmdu.2022.001267

Chen, J., Cui, D., Luo, R., et al. (2019). Multi-channel Human-Machine Interaction Technology for Military Command Systems. Command Control and Simulation, 41(04), 110–113.

Gupta, H. P., Chudgar, H. S., Mukherjee, S., et al. (2016). A continuous hand gestures recognition technique for human-machine interaction using accelerometer and gyroscope sensors. IEEE Sensors Journal, 16(16), 6425–6432.

Liu, F., Han, J., Qi, J., et al. (2021). Research and Application Progress of Intelligent Wearable Devices.

Luo B, Sun Y, Li G, et al. Decomposition algorithm for depth image of human health posture based on brain health. Neural Computing and Applications, 2020, 32(10): 6327–6342.

Nie, G. (2020). Research on human body gesture detection and recognition methods [Master's thesis, Nanjing University of Posts and Telecommunications]. https://doi.org/10.27251/d.cnki.gnjdc.2019.000844

Tan, C., Sun, Y., Li, G., et al. (2020). Research on gesture recognition of smart data fusion features in the IoT. Neural Computing and Applications, 32(22), 16917–16929.

Wu, Y. (2014). Research and application of gesture recognition technology based on depth image information [Master's thesis, Northeastern University].

Xie, J., & Zhang, L. (2015). Intelligent Wearable Devices and Their Applications. China Medical Device Information, 3, 18–23.

You, F. (2022). Gesture estimation based on key point detection [Master's thesis, Harbin Institute of Technology]. https://doi.org/10.27061/d.cnki.ghgdu.2022.002175

Yuan, G., Bing, R., Liu, X., et al. (2022). Spatial-Temporal Graph Neural Network based Hand Gesture Recognition. Acta Electonica Sinica, 50(4), 921–931.

Yuan, R. (2019). Research on gesture recognition method based on convolutional neural network [Master's thesis, Guangxi Normal University].

Zhang, W., Lin, Z., Cheng, J., et al. (2021). A survey of dynamic gesture understanding and interaction. Journal of Software, 32(10), 3051–3067. doi: 10.13328/j.cnki.jos.006217.

Zhao, F. (2016). Gesture recognition and application based on Kinect skeleton information [Master's thesis, Taiyuan University of Technology].