Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/36/20230430

Hand gesture recognition in natural human-computer interaction

Chang Ge[1,3], Jianhua Min[2]

[1]Columbia International College, Hamilton, Canada
[2]American Straight Academy, USA

[3]2023081047FBK@cic.care

Abstract. This paper introduces the definition of Gesture recognition. First the article gives a precise definition of gesture recognition and explains the difference between gestures and postures, then it reveals technical difficulties of Gesture recognition. These technical difficulties include four aspects. After analyzing the technology and methods of Gesture recognition and the technical difficulties, it concretely expounds gesture recognition process based on data gloves, which is widely studied before and it also introduces the computer vision based research achievement which is currently becoming a research hotspot in this field. Then it also takes Kinect and HoloLens as examples to introduce specific practical cases of gesture recognition in wearable devices. Also it outlines the application of gesture recognition technology in human-computer interaction which include but not limited to smart terminal, Game control, Robot Control, Clinical and Health, Smart home, Sign Language Recognition, Vehicle system, Interactive entertainment. Finally it reach the conclusion that the biggest challenge researchers meet is to build a powerful framework to overcome the common problems with fewer constraints to provide reliable results. And sometimes researchers need to combine multiple methods for different complex environments.

Keywords: Human–computer Interaction, Gesture Recognition, Data Glove, Computer Vision.

1. Introduction

Human-computer interaction refers to the way humans interact with computers or other digital devices. Gesture is a intuitive and easy way to learn human-computer interaction method. By using human hands as the input device of the computer, intermediate media is not required for communication between humans and computers. Data glove based and computer vision based technology are the two categories of gesture recognition technology

Data glove refers to a hardware device that collects hand movement data through built-in sensors in the glove. For Gesture recognition based on sensor data, general data processing methods include template matching method and neural network algorithm [1, 2]. The computer vision based approach offers non-contact communication way between humans and computers. The current computer vision recognition research is divided into 7 areas: Skeleton-Based Recognition, Motion-Based Recognition, Depth-Based Recognition, 3D Model-Based Recognition, Appearance-Based Recognition, Color-Based Recognition, Deep-Learning Based Recognition [3].

DOI: 10.54254/2755-2721/36/20230430

In the past decade, a large number of research results based on computer vision technology have been published. A study by He Maolin He proposes a gesture recognition method combining Hu moments and support vector machines, and designs experiments to verify ten digital gestures from 0 to 9. This paper first discusses some commonly used image preprocessing methods, and then introduces the feature extraction of gesture images and the theory of support vector machines [4].

A study by Luo Guoqiang, Li Jiahua, Zuo Wentao [5] researches the main gesture recognition techniques including template matching method, dynamic gesture recognition method using SVM and dynamic gesture recognition method using DTW.

A recognition system study presented by Khan [6], his research involves the recognition process including the gesture feature extraction, then the gesture classification, also the discussion of application area.

Li Wensheng, Xie Mei, Deng Chunjian [7] proposes an efficient multi-target detection and tracking method based on HSV color space, which realizes real-time detection and tracking of multiple fingertip targets through the camera; defines a set of dynamic gesture models based on fingertip movement trajectories, and proposes a dynamic gesture recognition method.

Wearable devices Kinect and HoloLens provide a very good example of commercial human-computer interaction, and have made significant progress in practice [8, 9]. Using computers to recognize gestures provides a more natural human-machine interface. However, since all kinds of computer vision have their limitations, the current challenge is to develop reliable and robust algorithms to solve common problems and obtain reliable results with the help of camera sensors with specific characteristics.

2. Technical Difficulties of Gesture Recognition

In the Chinese dictionary, gesture refers to the posture of the hand, and specifically refers to the specific position and body position changes when a person uses the arm. Gestures are considered to be the earliest and widely used communication method. Gesture recognition is the subject of recognizing human gestures through mathematical algorithms and it is an important part of human-computer interaction, and its research and development affect the naturalness and flexibility of human-computer interaction.

At present, the following three aspects are the difficulties in realizing natural gesture detection and recognition mainly:

The accuracy of gesture recognition is affected by many factors, such as lighting, background, gesture speed, etc. How to eliminate the influence of these factors is one of the problems that gesture recognition technology needs to solve

Gesture recognition technology will process huge gesture information, and the processing and analysis of these gesture information requires corresponding computing resources, and to improve computing efficiency is one of the problems that gesture recognition technology needs to solve.

(3) Gestures often have complex and multiple meanings, and it is difficult for a single method to accurately interpret the specific connotations of human hands and then realize the final recognition, so multiple methods need to be integrated.

Others include lighting changes, occlusion effects, and complex background reality limitations. For example, human hand movements are flexible and complex. There are not only directly visible and clear hand shapes, but also movements such as clenching fists, crossing fingers, overlapping left and right hands, etc. that greatly obscure joint points. How to accurately identify these occluded parts is difficult.

3. Gesture recognition technology and method

The most prominent research methods in gesture recognition include glove-based recognition(Contact recognition) and computer vision based recognition(non-contact). The recognition method using gloves mainly uses optical fiber to obtain the joint position and bending degree of the palm and fingers and model them. The computer vision-based recognition method refers to the acquisition of gesture vision images from the camera for a series of algorithmic processing to perform recognition and obtain results.

DOI: 10.54254/2755-2721/36/20230430

3.1. Gesture recognition process

Gesture recognition is a process of searching for human gestures, recognizing their different styles, and converting them into machine commands semantically.

The general process of gesture recognition as show in Figure 1.

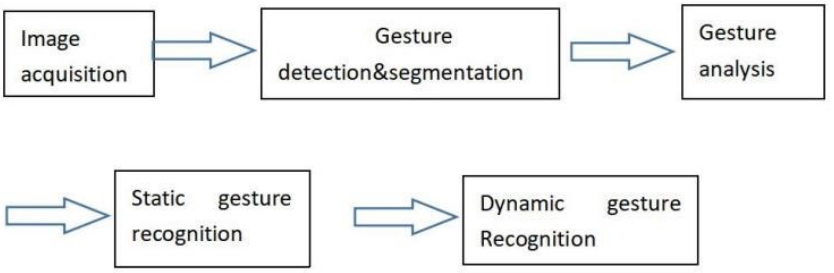

Figure 1. Process of gesture recognition.

In terms of specific implementation, vision-based gesture recognition technology usually includes the following steps:

1. Image collection: use devices such as cameras or depth sensors to collect hand images or three-dimensional information.
2. Preprocessing: Preprocessing the collected images or information, such as filtering, noise reduction, cropping, etc.
3. Feature extraction: Extract meaningful features from preprocessed images or information, such as the position of finger joints, the direction of the palm, the degree of bending of fingers, etc.
4. Classification and recognition: The above results are used as input, and the classifier is used for classification and recognition to obtain the type of gesture.

3.2. Based on data gloves

3.2.1. Basic conceptions. The data glove is a multi-mode virtual reality hardware. Through software programming, it can perform actions such as grabbing, moving, and rotating objects in the virtual scene.

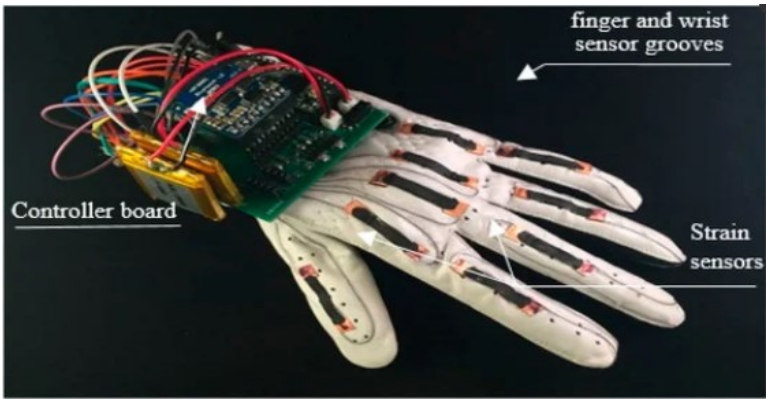

Figure 2. Data glove [10].

The data glove is mainly composed of an optical fiber fixed on the elastic fiber glove and a belt-type three-dimensional tracking sensor. There is an optical fiber ring at the joint of each finger. The optical fiber is bundled into several paths by plastic tubes to adapt to the bending movement of the finger.

DOI: 10.54254/2755-2721/36/20230430

When the hand is in a straight state and the optical fiber in the glove is also straightened, since the refractive index of the core wire is greater than that of the cladding, total reflection occurs, and all incident light can be transmitted to the other end of the optical fiber, so the amount of light transmitted is not reduced; When the finger bends and causes the optical fiber at the joint to deform, the refractive index of the optical fiber cladding will change, which does not meet the conditions of total reflection, and the light will be refracted. Only part of the total reflection light can be transmitted to the other end of the optical fiber. According to the intensity of light reaching the other end, The curvature of the finger can be determined.

Tactile feedback data glove is formed by adding tactile feedback unit on fingertips and palms on the basis of data glove. During the the process interaction, when the user's hand touches an object in the virtual environment (environment created in the computer by means of graphics), it can simulate the tactile feedback of vibration at the corresponding position of the human hand , to produce the feeling of interacting with the virtual world, that is, to obtain the so-called "immersion"; tactile sensors are usually made of micro switches, conductive rubber, carbon sponge, carbon fiber, pneumatic reset devices and other types.

When wearing a data glove with force feedback to grab a virtual object, the glove can generate a force that moves the human hand outward. The magnitude and direction of this force must be the same as the magnitude and direction of the force generated by the real object.

3.2.2. Data processing methods. From the technical realization of gesture recognition, common gesture recognition methods mainly include: template matching method and neural network method.The template matching method regards the gesture action as a sequence composed of static gesture images, and then compares the gesture template sequence to be recognized with the known gesture template sequence to recognize the gesture.

As for neural network method, general three-layer structure shown as Figure 3:

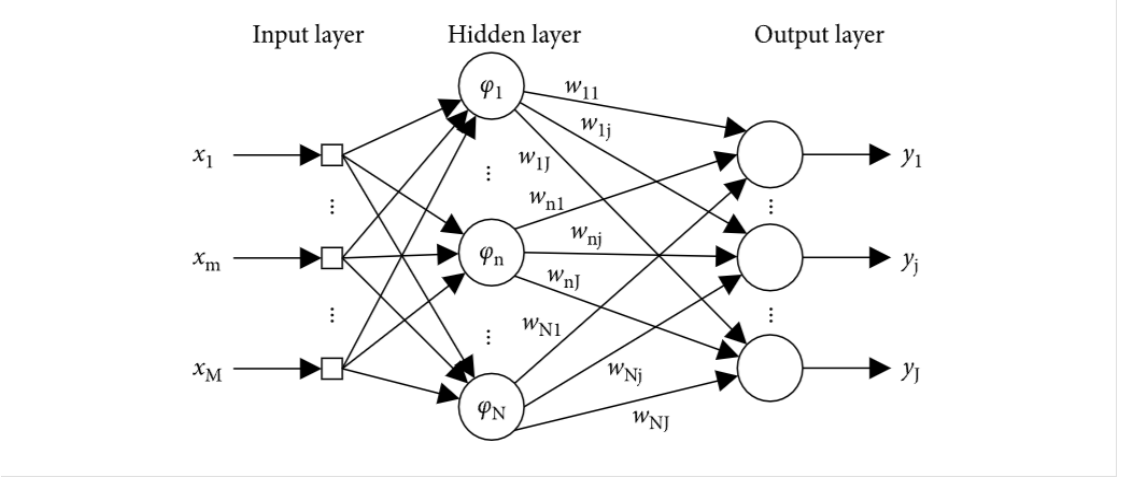

Figure 3. Three-layer structure neural network.

The basic process of the neural network Gesture recognition technology is as follows: first, obtain the image and action data of the gesture, then process and extract the features of these data, then learn how to recognize the gesture through the training model, and finally conduct Gesture recognition and control through the model. The performance of neural network Gesture recognition technology is closely related to the feature extraction method, neural network model and training data used. How to select and adjust these parameters has a great impact on the performance.

3.2.3. Method limitations. The data glove approach has limitations, it requires users to physically connect to the devices, which bring the convenience to the users. Also The data gloves have the

DOI: 10.54254/2755-2721/36/20230430

following disadvantages: (1) Practical and economic aspects: the accuracy of the packaged sensors used on the market is greatly limited, and it is difficult to upgrade; (2) Structure: Data gloves on the market have complex structures and high production costs.

3.3. Computer vision

3.3.1. Basic concepts. Computer Vision refers to the process of simulating human vision through visual information such as digital images or videos, so as to achieve the purpose of object understanding, recognition, classification, tracking, reconstruction, etc. Computer vision involves the automatic extraction, analysis and understanding of useful information from a single image or sequence of images. It involves the development of theoretical and algorithmic foundations for automatic visual understanding.

3.3.2. Research classification and trends. The current computer vision recognition research is divided into 7 areas:Skeleton-Based Recognition, Motion-Based Recognition, Depth-Based Recognition, 3D Model-Based Recognition, Appearance-Based Recognition, Color-Based Recognition, Deep-Learning Based Recognition.

With the rapidly development of technologies such as deep learning,machine learning and big data, computer vision technology has developed rapidly, and its application scenarios have also expanded. It has become one of the most popular artificial intelligence technology.

Further development of deep learning technology especially convolutional neural network has strong application value in image recognition and other aspects. In the future, deep learning technology will be further developed and will be applied to a wider range of scenarios, such as face recognition, smart homes, smart cars, etc.

Technological development of fusing Multimodal data: Multimodal data includes images, videos, voice, text, etc. Fusion of these data can improve the precision and accuracy of computer vision processing. In the future, computer vision technology will pay more attention to the processing and utilization of multimodal data for optimization of robustness and accuracy of the algorithm.

Application of edge computing: edge computing refers to pushing computing and storage resources closer to the data source, such as sensors and terminal devices. Computer vision technology requires a lot of computing and storage resources, and edge computing can improve computing efficiency, reduce latency and energy consumption, so computer vision technology will be more applied to edge computing scenarios in the future.

4. Business case

Microsoft Kinect, also known as Microsoft motion sensor, is a hardware device specially used for Xbox360 game console released by Microsoft in 2010. It includes a camera, an infrared blaster and a set of microphones. Capture the space in front of the host through infrared emitters and cameras, so as to realize player posture detection, action recognition, voice recognition and other functions.

4.1. Microsoft Kinect introduction

Microsoft Kinect does not require any physical contact, and realizes human-computer interaction control by recognizing human actions and gestures. At present, kinect has been used in many fields such as activity recognition, gesture control, and smart home. Microsoft Kinect, in addition to being able to ingest RGB color information, can also use infrared light to provide depth information. This provides great convenience for gesture detection, because the infrared depth information is much more distinguishable from the background than the color information and motion information, so that most of the work of gesture detection can be done directly by hardware.

DOI: 10.54254/2755-2721/36/20230430

4.2. Microsoft HoloLens case

On this basis, Microsoft's subsequent wearable device, HoloLens, can complete some commands and operations in a mixed reality environment through gestures. The interaction method of HoloLens is to use gaze to locate (similar to where the mouse cursor moves), and then use gestures or sounds to manipulate any target being located (similar to a mouse click). Although gestures cannot accurately locate in space, their advantage is that they allow users to quickly operate HoloLens headsets without the need for other equipment.

The positioning mechanism of gesture interaction is gaze. The combination of gaze and air tap constitutes the interaction between gaze and action. HoloLens is currently able to recognize two core gestures - Air tap and Home gestures. These two core gestures are the most basic spatial input units that developers can obtain, laying the foundation for users to perform more diverse subsequent operations.

Air tap is a gesture that requires the hand to be upright and clicked, similar to clicking with a mouse. This gesture is equivalent to a "click" in many HoloLens applications, where users can click through the Air tap after determining the target through staring. As long as you learn this gesture, it can be widely used in various applications. As shown in Figure 4.

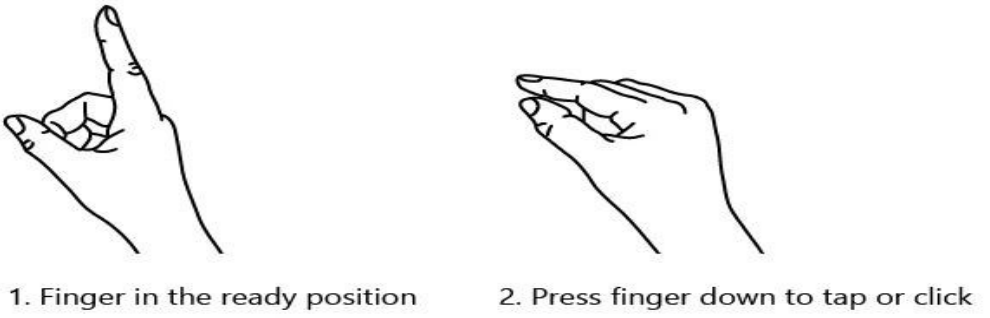

Figure 4. Gesture in different states.

The gesture of Bloom in HoloLens represents "home" and does not give it any other meaning. This gesture in the system represents returning to the start interface, which has the same meaning as the Windows start button or the Xbox homepage button. As shown in Figure 5.

Figure 5. Gesture Air TAP and Bloom.

5. Application of gesture recognition

Gesture recognition technology has been used, including but not limited to real-world scenarios.

DOI: 10.54254/2755-2721/36/20230430

1) Smart terminal: Gesture recognition technology can help users perform various operations on mobile phones or tablets, such as sliding, zooming, and rotating; 2) Game control: Gesture recognition technology allows players to use gestures in the game to control the character's walking, jumping, etc. 3) Smart home: Gesture recognition technology allows users to control home devices through gestures, such as turning on the TV and adjusting lights. For example, in a smart home system, all the lights can be turned off by simply waving your arm down. 4) Vehicle system: Gesture recognition technology allows the driver to use gestures in the car to control various functions, such as adjusting volume, changing music, etc. For example, in some car systems, you can switch to the next song just by moving your palm to the right. 5) Interactive entertainment: In these scenes of live video, people often want to interact with the person on the opposite side of the screen, and combine user gestures (such as likes) to add corresponding stickers or special effects in real time. 6) Sign Language: Gesture recognition technology determines the meaning expressed by hand movements through the analysis and recognition of human hand movements. At present, gesture recognition technology has achieved high accuracy and can recognize most of the hand movements.Sign language recognition technology has been used in the field of education. The Educational Association for the Deaf is working to provide a more equal and inclusive learning environment for deaf students through the use of sign language recognition technology. In addition, in some special schools, sign language recognition technology can also be used to assist teachers in teaching and students in learning.7) Clinical and Health:gesture recognition technology is also used in the medical field. For example, in some hospitals, doctors can use sign language recognition technology to communicate with hearing-impaired patients. In addition, in some rehabilitation centers, deaf patients can also use sign language recognition technology to obtain better rehabilitation services.

6. Conclusion

In gesture recognition research, to design a powerful framework that overcomes problems with fewer constraints and provide reliable results is the biggest challenge researchers meet .

Contact recognition and non-contact recognition technology are the two categories in Gesture recognition technology. Due to the many limitations of data glove technology, the research focus has gradually shifted to computer vision, and great progress has been made. Computer vision recognition research can be subdivided into several fields, each corresponding to a different scenario, but all have one thing in common: a lot of effort is required to develop reliable and powerful algorithms, with the help of sensors with specific characteristics to solve common problems and obtain reliable results. Each of the above techniques has its advantages and disadvantages, depending on the specific scenario in which it is used.In real life, gestures are usually in complex environments, so multiple methods need to combined to accrue more recognition accuracy.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Guo K, Zhang S, Zhao S, et al 2021. Design and manufacture of data gloves for rehabilitation training and gesture recognition based on flexible sensors. Journal of Healthcare Engineering, Access 13.

[2] Li Y, Yang L, He Z, et al 2022. Low‐cost data glove based on deep‐learning‐enhanced flexible multiwalled carbon nanotube sensors for real‐time gesture recognition Advanced Intelligent Systems, 4(11): 2200128.

[3] Oudah M, Al-Naji A, Chahl J 2020. Hand gesture recognition based on computer vision: a review of techniques. journal of Imaging, 6(8), 73.

[4] He Maolin. (2016). Research and Implementation of Gesture Recognition Algorithms Based on Computer Vision. (Doctoral dissertation, University of Electronic Science and Technology of China).

DOI: 10.54254/2755-2721/36/20230430

[5] Luo Guoqiang, Li Jiahua, Zuo Wentao, Fang Bin. Research on Steps and Methods of Gesture Recognition Based on Computer Vision Technology
Wireless Internet Technology Magazine Agency 2020, Vol. 17 Issue (3): 148-149.doi: 10.0002/1672-6944-1718

[6] Khan R Z, Ibraheem N A 2012. Hand gesture recognition: a literature review. International journal of artificial Intelligence & Applications, 3(4), 161.

[7] Li Wensheng, Xie Mei, Deng Chunjian. A dynamic multi-point gesture recognition method based on machine vision [J]. Computer Engineering and Design, 2012 (5): 1988-1992.

[8] Tang M 2011. Recognizing hand gestures with microsoft's kinect. Palo Alto: Department of Electrical Engineering of Stanford University: [sn], 23.

[9] Pal D H, Kakade S M 2016. Dynamic hand gesture recognition using kinect sensor. In 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC) 448-453.

[10] Murthy G R S, Jadon R S 2009 . A review of vision based hand gestures recognition. International Journal of Information Technology and Knowledge Management 2(2): 405-410.