Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/4/20230503

Review of the development of gesture recognition technology in recent years

Xuehan Gao

Mathematical Sciences, University of Southampton, Southampton, SO17 1BJ, UK

xg1n21@soton.ac.uk

Abstract. Currently, gesture recognition is the most rapidly advancing field in processing images and artificial technology. Gesture processing entails a procedure whereby gestures of the human body are recognized and used to coordinate PCs and other intelligent devices. This research will broadly outline human and PC coordination through gesture recognition technology over the past few years. Gesture recognition develops a simplified communication path between humans and PCs called Human-Computer Interaction (HCI). This study seeks to expose the significance of gesture recognition technology in recent years. In addition, this research will outlay the various methods in which gesture recognition technology can be applied in different perspectives, namely in education, transport, healthcare, and entertainment. This research will also elaborate on the pros and cons of every application process in each aspect. Lastly, this research will give an insight into the negative factors that hinder the development of gesture recognition technology. The research of gesture recognition mainly focuses on the visual aspect, and gesture provides an easy to understand recognition method. This makes gesture recognition products come into being and are widely used in the world. Now people do not need to touch anything. As long as they can wave their hands in the air, they can realize a series of complex operations that can be completed by touching the screen with the mouse and keyboard in the past.

Keywords: Gesture Recognition, Human-Computer Interaction, Pattern Recognition.

1. Introduction

Gesture recognition is a functioning study field that entails the integration of the gestural channel in (HCI) Human-Computer Interaction (Sun et al., 2020). The gesture recognition technology functions under a virtual ecosystem control which allows the signing in of language translation, music production, and even robot remote controls. Human recognition gestures come in a broad context of pattern recognition (Zhao, 2020). In this context, systems entail two procedures: the decision process and representation. The representation procedure involves converting raw numerical data into a mode modified to the decision procedure, which then classifies the in Research Background and Current Situation.

With the rapid development of computer science, people interact with computers more and more. Most of the interaction between people and computers is carried out by mouse, keyboard and touch device. However, the interaction mode through hardware is not easy for people to understand and limits the speed of human-computer interaction. People put forward higher requirements for human-

DOI: 10.54254/2755-2721/4/20230503

computer interaction between the real world and the virtual world, and are committed to studying a convenient way that is more in line with human natural interaction.

The rise of artificial intelligence technology has spawned many new human-computer interaction methods making the human-computer interaction more diversified. In virtual scene, gesture recognition can transform human gesture semantics and establish a more natural human-computer interaction mode. In recent years, the research of gesture recognition mainly focuses on the visual aspect, and gesture provides an easy to understand recognition method. This makes gesture recognition products come into being and are widely used in the world. Now people do not need to touch anything. As long as they can wave their hands in the air, they can realize a series of complex operations that can be completed by touching the screen with the mouse and keyboard in the past.

2. Gesture Recognition based on Image

Gesture recognition technology structures inherit this system and are involved in two other procedures: The acquisition activity entails converting physical gestures into numerical information (Yasen & Jusoh, 2019). The other procedure is the interpretation activity which offers the meaning of all symbols derived from the decision procedure. Any hand gesture comprises four aspects: movement, hand configuration, location, and orientation. A primary classification involving gesture recognition technology can also be achieved by the separation of dynamic gestures, which are a classification of hand postures, and static gestures, which are referred to as hand postures (Yasen & Jusoh, 2019).

3. Main Modes of Gesture Recognition

3.1. Input-basic processing flow-output

Gesture input is an aspect in charge of receiving gesture motions with the transducer's purpose offered by the leap gesture control. It turns the skeletal information, attained in frames, into 2-D aspect vectors and then conveys it to the inner erudition neural net substructure. A leap motion control functions near top level precision trails frame rate and outlays distinct locations and motions. It puts to affect two optical devices and three ultraviolet illuminations. The optical devices are directed alongside the Y axis, ascending when the regulator is in its typical functioning position and has a field vision of around 150 degrees. The Leap Motion Control system's operative range encompasses about 20 to 610 millimeters directly above the device. The data that leap motion Control conveys is the hand skeleton model it sees in the form of frames. The skeletal model comprises entities (like tools or fingers) positions comparative to leap motion controls derivation point. The leap motion system engages a right-tendered Cartesian coordinate scheme.

3.2. Overview of gesture recognition based on interaction mode

With the upsurge in gesture recognition technology, the need to develop an effective human-computer interaction system technology has been a relentless journey in the field of research. There are two significant categories of gesture acquisition structures: vision and device-based (Li et al., 2018). In a deeper look into machine-based systems, the achievement of gestures is developed by applying various gesture interaction technology procedures in virtual reality. In vision-based systems, the motions are taken by a camera. This section seeks to outlay the Input-Basic Processing Flow-Output in gesture interaction technology and elaborates on the classification of input appliances used in gesture recognition appliances.

DOI: 10.54254/2755-2721/4/20230503

Table 1. List of datasets used by referenced papers.

Author	Dataset
Abigail	• Interaction in Medical Imaging • Enus research project • Observation method
John	• Fast Hand Gesture Recognition Using representative Frames • International Conference on Digital Image Computing • Techniques and Applications
Rand	• The Sony PlayStation II EyeToy • Virtual Reality for in Rehabilitations • Physical observation
Ma T	• Kinect-based Gesture Recognition • International Conference on Signal Processing • Surveys and data computing
Shweta K	• Kinect-based Gesture Recognition • International Conference on Signal Processing • Surveys and data computing
Sun	• Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images • IET Image Processing • Physical observation
Yasen	• Hand gesture recognition techniques, challenges and applications • Computer Science • Surveys and data computing
Li	• Dynamic gesture recognition in the internet of things • Computer science • Surveys and data computing
Duan	• Gesture recognition based on multi-modal feature weight • Concurrency and Computation • Surveys and data computing
Tan	• Gesture recognition of smart data fusion features in the IoT • Neural Computing and Applications
Yin	• Gesture recognition technology of data glove based on joint algorithm • International Conference on Mechanical, Electronic, Control and Automation Engineering • Surveys and physical observation
Zhao	• Machine learning and gesture recognition • International Conference on Computing and Data Science • Surveys and data computing

3.3. Gesture recognition based on convolution network and recurrent network

Convolution network is to input an image. For example, if there is an apple in this image, this image can be input into the network. After training, a label can come out of the network. It shows that this picture is an apple. The target detection, as an example, can monitor the position of the apple, and can draw the specific position of the apple. If it is used as gesture recognition, only inputting a single image into convolution network is not enough, because only by inputting a single image into the network, it is not enough to recognize a complete gesture. If the system can recognize a gesture, it needs multiple frames. Multiple frames can form a continuous action. An action can be entered into

DOI: 10.54254/2755-2721/4/20230503

the system by inputting multiple pictures, so that it can generate a label. The method we focus on is to input this continuous multi-frame image into the system, encode and process the image. Recurrent network is the most commonly used method to adapt to the situation of multiple frames.

The Figure 1 shows the process of gesture recognition by recurrent network. Recurrent network can consider data of a time series, which are data with multiple sampling points in time. After the data comes into the system, the characteristics of changes generated in a sequence are output through a recurrent network. This is a many-to-one process. A label from the system tells what gesture it is. If the recurrent network is used, the activation state after the previous frame comes in needs to be retained. The next frame comes in and forms a sequence. After the next frame comes in, it will also refer to the activation state after the previous frame comes in. It is equivalent to combining the different activation states after the whole sequence enters the network to form a feature. After a feature in this time dimension is output, a label is finally output. This is how gesture recognition been achieved by recurrent network. On the other hand, it is also possible to use convolution network. With separable convolution layer, this is a vertical one, and multiple frames can be compressed into one. You can input multiple frames into the convolution network as multiple channels, and then output a label to express a gesture.

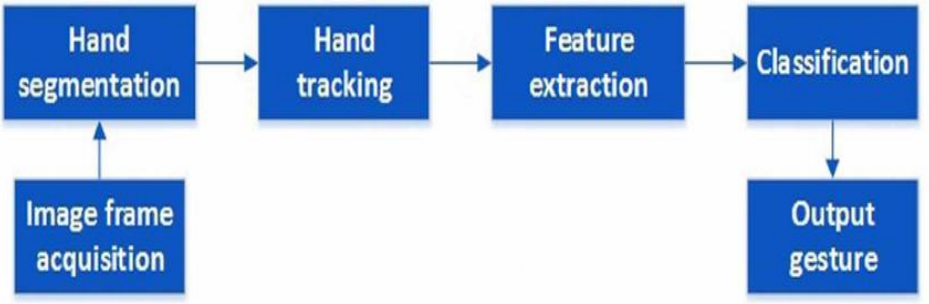

Figure 1. The flow chart of gesture recognition.

3.4. Classification of gesture recognition based on interaction mode

Image verification is rapidly becoming a vital step in most current problem-solving structures (Duan et al., 2021). Gestures are classified into direct contact, non-direct contact and media gestures. Media gestures comprise a solitary point input device through physical contact using devices like a trackball, mouse, or joystick to transmit the assorted data into the computer systems (Duan et al., 2021). On the other hand, direct contact gestures are gesture procedures outlaid directly on the input appliance through a physical item tool or a body part. When the operator gets into contact with the device, such as the screen, the gesture interaction mode is activated, and the touch screen response is the immediate response of interaction. In a deeper perspective, direct contact gestures comprise single and several touch gestures. Single touch gestures operate by using a single point of focus using a particular pointing input appliance like an intelligent pen or a mouse (Yin et al., 2018). A smart pen is a single-touch gesture appliance that identifies the movement of a pen and uses a symbol for the express intention. Conducting commands without physical contact with the appliances is known as non-direct contact gestures. An example of a non-direct contact gesture appliance is a plurality of light emitters.

4. Main Modes of Gesture Recognition

Gesture recognition use sensors to interpret commands. Gesture recognition is an exciting experience for it is flexible which enables the user to handle various multiple points of input and also define multiple aspects at once. Gestures also express the natural mode of communication. In addition gesture recognition technology is immediate and decisive when compared to traditional buttons, gesture technology far much reliable since they do not interrupt the user's activity by forcing them to move to the command location. More over Gesture technology can be performed directly from the operator’s

DOI: 10.54254/2755-2721/4/20230503

location. The most interesting aspect about gesture technology is that it is enjoyable to operate with when compared to other inputs like initiating command using a mouse input, the direct touch input is fast, and studies have revealed that most individuals prefer direct touch devices due to their simplified nature.

Designers must carefully consider if gesture technology poses a risk or improves consumer experience. For instance some gesture recognition devices have accuracy issues and delivering wrong commands can turn out tragic to the consumers. In addition Discoverability is a huge disadvantage in Gesture recognition technology. Discoverability is a significant limitation facing the use of intelligent devices. Some gesture recognition devices are not self-explanatory, nor does self-revealing make it a challenge for them to be discovered by a less skilled individual. Fatigue: Despite gestures being enjoyable to use, they also cause fatigue since they are pretty engaging since some gesture devices entail complex activities for a long time compared to other interaction procedures.

5. Conclusion

The vigorous development of artificial intelligence has brought opportunities and challenges to human-computer interaction. Providing a natural and effective human-computer interaction interface has always been the purpose of human-computer interaction research. The future development of gesture recognition should also focus on this point. In the future development, it is necessary to complement spoken language, gesture and face in the virtual environment. Because there are two ways of gesture: 3D based modeling and image-based modeling. 3D based modeling has strong expression ability, and can express almost all hand shapes, but it lacks efficiency. The latter is simple and efficient, but lacks universality. Gestures consist of a series of hand shapes over a period of time.

References

[1] Abigail S. Touch-less Interaction in Medical Imaging[R/OL]. https://www.microsoft.com/enus/research/project/touchless-interaction-in-medical-imaging

[2] John V, Boyali A, Mita S, et al. Deep Learning-Based Fast Hand Gesture Recognition Using Representative Frames[C]//2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, QLD, 2016: 1-8.

[3] Rand, Debbie, Kizony, Rachel, Weiss, Patrice (Tamar) L. The Sony PlayStation II EyeToy: LowCost Virtual Reality for Use in Rehabilitation[J]. Journal of Neurologic Physical Therapy, 2008, 32(4): 155-163

[4] Ma T, Guo M. Research on Kinect-based Gesture Recognition[C]//2019 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Dalian, China, 2019: 1-5.

[5] Shweta K, Yewale, Pankaj K. Bharne, <Artificial Neutral Network Approach Gesture Recognition>, <International Journal of Engineering Science and Technology> Vol. 3 No. 4 April 2011

[6] Sun, Y., Weng, Y., Luo, B., Li, G., Tao, B., Jiang, D., & Chen, D. (2020). Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images. *IET Image Processing*, *14*(15), 3662-3668.

[7] Yasen, M., & Jusoh, S. (2019). A systematic review on hand gesture recognition techniques, challenges and applications. *PeerJ Computer Science*, *5*, e218.

[8] Li, G., Wu, H., Jiang, G., Xu, S., & Liu, H. (2018). Dynamic gesture recognition in the internet of things. *IEEE Access*, *7*, 23713-23724.

[9] Duan, H., Sun, Y., Cheng, W., Jiang, D., Yun, J., Liu, Y., ... & Zhou, D. (2021). Gesture recognition based on multi-modal feature weight. *Concurrency and Computation: Practice and Experience*, *33*(5), e5991.

[10] Tan, C., Sun, Y., Li, G., Jiang, G., Chen, D., & Liu, H. (2020). Research on gesture recognition of smart data fusion features in the IoT. *Neural Computing and Applications*, *32*(22), 16917-16929.

DOI: 10.54254/2755-2721/4/20230503

[11] Yin, S., Yang, J., Qu, Y., Liu, W., Guo, Y., Liu, H., & Wei, D. (2018, March). Research on gesture recognition technology of data glove based on joint algorithm. In *2018 International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018)* (pp. 41-50). Atlantis Press.

[12] Zhao, P. (2020, August). A review on machine learning and gesture recognition. In *2020 International Conference on Computing and Data Science (CDS)* (pp. 425-428). IEEE.