Advances in Vision-Based Gesture Recognition and Its Diverse Applications

Jidong Liu [1], Yehao Tan [2] and Zhaohe Wu[3,*]

[1] School of Computer Science and Technology, Jilin University of Finance and Economics, Jilin, China

[2] School of Computer Science and Technology, Southwest Petroleum University, Nanchong, China

[3] School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou, China

*2022337621027@mails.zstu.edu.cn

Abstract. With the rapid development of science and technology, gesture recognition has emerged as one of the most important methods of human-computer interaction. Among the various approaches to gesture recognition, vision-based methods have proven to be the most mature and widely adopted. Vision-based gesture recognition leverages advanced computer vision techniques and machine learning algorithms to interpret human gestures from visual data captured by cameras. This method allows for intuitive and natural interactions between humans and computers, enabling applications in a wide range of fields. From virtual reality and gaming to sign language interpretation and smart home controls, vision-based gesture recognition offers significant potential for enhancing user experiences. Its ability to provide touchless control and seamless interaction makes it a crucial component in the advancement of human-computer interfaces. As technology continues to evolve, vision-based gesture recognition is expected to play an increasingly pivotal role in creating more immersive and responsive interactive systems.

Keywords: Computer Vision Gesture Recognition Overview.

1 Introduction

Gestures are a crucial form of human communication alongside language, serving as a direct and natural means of human-computer interaction (HCI) [1]. As a non-verbal communication tool, gestures play an essential role in daily life, directly conveying emotions and attitudes, such as waving goodbye or shaking hands to show friendship [2]. Statistics indicate that everyday gestures can help transmit hundreds of thousands of different messages, which are easy to understand yet rich in emotional content. The advancement of gesture recognition technology has enabled machines to understand human gestures, facilitating applications in various life scenarios, including human-computer interaction, sign language communication, and intelligent medical care [3]. Initially, gesture recognition technology relied on data gloves and optical

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_22

markers to detect gestures, but this method required specific equipment. Recently, the development of computer vision and pattern recognition technologies, coupled with machine learning and deep learning, has significantly advanced the field [4]. However, computer vision-based gesture recognition still faces challenges due to the complexity of hand structures and visual backgrounds, the diversity of gesture semantics, and poor image quality caused by inadequate lighting and dynamic motion [5].

2 Current status of research at home and abroad

With the transformation of deep learning in the field of computer vision, gesture recognition technology has attracted a lot of attention from researchers around the world [6, 7]. At present, scholars at home and abroad are committed to further improving its diversity, accuracy and real-time performance on the basis of existing gesture recognition technology, and improving its ability to obtain accurate gesture patterns in complex contexts [8]. Gesture recognition technology can be divided into static gestures and dynamic gestures, and gesture model recognition can be divided into template matching, data classification and deep learning gesture recognition algorithms.

2.1 Static gesture detection

Static gesture recognition is a gesture image at a certain time, and its recognition results are strongly related to the appearance characteristics of the hand in the image, such as location, outline, and texture [9].

2.2 Dynamic gesture detection

Dynamic gesture recognition is a series of images in a continuous period of time. The recognition results are related to the appearance of the hand in the image and the time series characteristics of the hand in the sequence [10]. Dynamic gestures are more diverse, expressive and practical than static gestures.

2.3 Gesture Recognition Algorithm Based on Template Matching

Template matching-based gesture recognition algorithm is mainly Dynamic Time Warping (DTW).

The DTW algorithm is a method proposed by Japanese scholars in the 1960s to measure the similarity between two time series of different lengths. When recognizing gestures using DTW algorithm, a series of reference templates are recorded in advance, and then the similarity between the gesture to be measured and the reference template is calculated. It has the characteristics of low demand for training samples and high accuracy, but it is also limited by high computational complexity and poor stability, especially in the case of complex gestures and large number of training samples. In order to improve the DTW algorithm and improve the efficiency of the algorithm, scholars have proposed many improved algorithms, such as: Fast DTW, Sparse DTW, LB_ Keogh, B_ Improved and LB_HUST algorithms.

2.4 Gesture Recognition Algorithm Based on Data Classification

Support Vector Machine (SVM) is based on traditional statistical learning theory. It shows many advantages in solving small sample, nonlinear and high-dimensional pattern recognition problems. Huang et al. proposed a gesture recognition method based on Gabor filter and SVM classifier. Dardas et al. used SIFT descriptor to extract the key points in the image, and mapped the key points into histogram vectors using K-means clustering algorithm and vector quantization algorithm.

Artificial Neural Network (ANN) is a class of information processing models inspired by biological neural networks, consisting of many interconnected parallel neurons. Different structures can be derived for different needs. A neural network algorithm based on improved probability is proposed. K-W test method is adopted to realize the characteristic selection of surface- Myoelectrogram Gestures. Propagation parameters are optimized by particle swarm optimization method.

Hasan et al. used the error backpropagation network (BPNN) to train features, and the recognition accuracy rate reached 70.83% and 86.38%, respectively. The optimization of BPNN by quantum particle swarm algorithm solves the problems of slow convergence speed, local minimization and network training failure, and improves recognition efficiency and stability.

Tusor et al. used Fuzzy Neural Network (FNN) to build a fuzzy gesture model and set 14 fuzzy eigenvalues between "large", "medium" and "small" to describe and distinguish different gestures. A self-growing and self-organized neural gas (SGONG) network is proposed, which extracts effective features by fitting the shape of the hand. This feature is insensitive to the scale and rotation of gestures. In addition, this network converges faster than other networks.

2.5 Gesture Recognition Algorithm Based on Deep Learning

Deep learning is unsupervised learning, which not only automatically extracts features from images, but also automatically learns higher-level features, overcoming the subjectivity and limitations of manual feature extraction. However, most existing methods may have problems with large models and slow execution. Make Skeleton-based Action Recognition Model Smaller, Faster and Better has proposed a two-characteristic two-motion network (DD-Net) for skeleton-based action recognition. By using a lightweight network architecture (i.e., 150,000 parameters) for better performance, the DHG-14 and DHG-18 datasets achieved accuracy of 94.6% and 93.9% respectively. Liu Jinfu et al. proposed a time decoupling graph convolution network (TD-GCN) that applies different adjacency matrices to skeletons from different frames.

3 Challenges and Solutions

3.1 Complex background

In visual gesture recognition, extracting the hand completely from the image is a key problem. At present, the depth threshold literature proposes that the object in a depth threshold space can be extracted by computer vision.

3.2 Dynamic capture

For dynamic gesture recognition, the precision, real-time, robustness, and how to accurately capture and recognize gestures in a series of fast motion recognition have become a great challenge. However, in the face of large-scale actions, gesture recognition still has certain challenges, and space for progress.

3.3 Visual noise and occlusion

During gesture recognition, visual noise in the image is caused by various environmental factors, such as light changes, background clutter, hand occlusion, etc. To solve these problems, image enhancement, background modeling, occlusion detection and other methods can be used to improve image quality and gesture visibility.

3.4 Diversity of gesture recognition

There is a certain diversity and variability in gesture categories, and there may be differences in gestures from person to person. To overcome this challenge, data augmentation techniques can be introduced to generate more training samples by rotating, scaling, twisting, and so on. In addition, methods such as transfer learning and cross-data set training can be used to improve the generalization ability of the model.

4. Application Analysis

4.1. Sign language communication

The number of deaf and mute people in the world is huge, and they can communicate with each other through sign language, but it is necessary and important to recognize the role and application of sign language communication through visual gesture. As summarized by the authors, the sign language recognition in China and other countries has been extracted and recognized using YOLO (You Only Look Once) model, CNN, CapsNet, etc. and trained using CSL-100 as a continuous sign language dataset. At present, sign language recognition has good results, but it is limited by the small amount of data in the dataset, and the interference of skin color, light, angle, recognition efficiency, and sign language and speech integrity need to be improved.

4.2 Smart Medical

With the advent of the information age, the computer industry is booming, medical care is gradually moving towards intelligence, information, and in the current computer has been widely used in medical and other fields, of which gesture recognition is one of the important. In is the application of visual gesture recognition in medicine. The author uses fusion and convolution features and key point features, and then combines RGB gestures to extract features.

It differs from traditional medicine in that it is not vivid and monotonous in teaching.

4.3 Virtual reality

Virtual reality is a technology that simulates the real world through a computer. In this field of virtual reality, virtual interactions can be made through gesture recognition. such as movement, rotation, etc. Through this interaction, it can effectively improve the user's sense of immersion and realism, give the user a close to the real experience, and make virtual reality more mature and close to reality. In this field of virtual reality, with the further development of gesture recognition technology, along with the development of other technologies, there will be a better experience in many sectors. In addition, in other sectors, such as education, military, etc., virtual reality technology with vision-based gesture recognition plays an expected role.

5. Conclusion

In conclusion, gesture recognition technology has emerged as a pivotal component in human-computer interaction, driven by advancements in science and technology. Among the various approaches, vision-based gesture recognition stands out due to its maturity and wide adoption. Leveraging advanced computer vision techniques and machine learning algorithms, this method allows for intuitive and natural interactions between humans and computers, enabling a plethora of applications across diverse fields such as virtual reality, gaming, sign language interpretation, and smart home controls. Vision-based gesture recognition technology has evolved significantly, transitioning from the use of data gloves and optical markers to sophisticated computer vision and pattern recognition methods. The development of machine learning and deep learning has further propelled this field, enabling more accurate and efficient gesture recognition even in complex and dynamic environments. Despite these advancements, challenges remain, such as handling complex backgrounds, dynamic capture, and the variability in gesture semantics and image quality. Current research both domestically and internationally focuses on enhancing the diversity, accuracy, and real-time performance of gesture recognition technologies. Scholars are exploring various algorithms, including template matching, data classification, and deep learning, to improve the efficiency and stability of these systems. Notable advancements have been made, such as the development of lightweight network architectures and innovative algorithms to address the limitations of existing methods.

Authors Contribution

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

References

1. Wu, W., Gao, W., Pang, B., et al.: Rapid Method for Chinese Sign Language Gesture Word Recognition. High Technology Communication (6), 23-27 (2001).
2. Sarma, D., Bhuyan, M. K.: Methods, Databases and Recent Advancements of Vision-Based Hand Gesture Recognition for HCI Systems: A Review. SN Computer Science 2(6), 436 (2021).
3. Al-Shamayleh, A. S., et al.: Systematic Literature Review on Vision-Based Gesture Recognition Techniques. Multimedia Tools and Applications 77, 28121-28184 (2018).
4. Oudah, M., Al-Naji, A., Chahl, J.: Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. Journal of Imaging 6(8), 73 (2020).
5. Franslin, N. M. F., Ng, G. W.: Vision-Based Dynamic Hand Gesture Recognition Techniques and Applications: A Review. In: Proceedings of the 8th International Conference on Computational Science and Technology (ICCST 2021), Labuan, Malaysia, August 28-29. Springer, Singapore (2022).
6. Zhao, R., et al.: Review of Human Gesture Recognition Based on Computer Vision Technology. In: 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Vol. 5. IEEE (2021).
7. Zhu, X., Xu, H., Zhao, Z., et al.: An Environmental Intrusion Detection Technology Based on WiFi. Wireless Personal Communications 119(2), 1425-1436 (2021).
8. Mohamed, N., Mustafa, M. B., Jomhari, N.: Review of the Hand Gesture Recognition System: Current Progress and Future Directions. IEEE Access 9, 157422-157436 (2021).
9. Jain, R., Karsh, R. K., Barbhuiya, A. A.: Literature Review of Vision-Based Dynamic Gesture Recognition Using Deep Learning Techniques. Concurrency and Computation: Practice and Experience 34(22), e7159 (2022).
10. Beddiar, D. R., et al.: Vision-Based Human Activity Recognition: A Survey. Multimedia Tools and Applications 79(41), 30509-30555 (2020).