A Comparative Analysis of Glove-Based and Image-Based Sign Language Recognition Systems

Junkang Rong
School of Computing, Civil Aviation Flight University of China, Guanghan, Sichuan, 618300, China

Keywords: Sign Language Recognition, Image Recognition, Glove-Based Sensors.

Abstract: As an important means of communication for deaf and hearing-impaired individuals, sign language possesses a unique linguistic system and mode of expression. However, because sign language is not widely spoken, the rate of standardized adoption is low, and there are significant differences in the sign languages used by various groups of deaf people. Consequently, sign language recognition technology plays a crucial role in facilitating barrier-free communication between deaf individuals and those with normal hearing. Sign language recognition systems that utilize glove-based sensors and image recognition have made significant advancements, thereby enabling more convenient communication. Glove-based sensors offer high precision by capturing detailed hand gestures. They are particularly effective in low-light conditions or when the signer is off-camera, providing a consistent recognition rate. On the other hand, image recognition systems excel in their non-intrusive nature, allowing for sign language interpretation without the need for the signer to wear any devices. They can process sign language in real-time and are ideal for video-based applications, making them suitable for inclusive social interactions and educational tools. This paper will review glove-based and image-based sign language recognition systems, covering their background, current research status, key technologies, and potential application prospects.

1 INTRODUCTION

According to the World Health Organization, there are more than 1 billion persons with disabilities globally, with the deaf accounting for 10 per cent of the disabled population. In China, eighty-two million people worldwide are classified as disabled, 20.54 million of whom have hearing impairments and 1.3 million of whom have speech impairments, according to the results of the Sixth National Population Census and the Second National Sample Survey of Persons with Disabilities. As the main means of communication for these people, sign language has a unique syntax, semantics and vocabulary system (Jones, 2021). However, since sign language is not a mass language, the prevalence of standardized sign language is very low, and there are dialectal phenomena in sign language used by different deaf groups, which poses a challenge to the development of sign language recognition technology.

Originating in the 1980s, sign language recognition technology has steadily received attention and research due to the ongoing advancements in computer technology. Early sign language recognition systems were mostly based on wearable devices such as data gloves, and realized the classification and recognition of sign language through multi-sensor fusion technology (Cheok, 2019). However, such systems have problems such as bulky equipment, high cost, and affecting the naturalness of human-computer interaction. In recent years, with the rise of computer vision and deep learning technologies, the sign language recognition system based on image recognition has gradually become the mainstream of research (Wadhawan, 2021). The system captures sign language images or videos through cameras, and realizes the automatic recognition of sign language using techniques such as image processing, feature extraction and classification algorithms.

This paper reviews advancements in sign language recognition, focusing on glove-based and image-based systems. It examines their technologies, evaluates their accuracy and convenience in facilitating communication for the deaf and hearing-impaired, and discusses their research status and applications.

Rong, J.
A Comparative Analysis of Glove-Based and Image-Based Sign Language Recognition Systems.
DOI: 10.5220/0013337000004558
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 1st International Conference on Modern Logistics and Supply Chain Management (MLSCM 2024)*, pages 408-412
ISBN: 978-989-758-738-2
Proceedings Copyright © 2025 by SCITEPRESS – Science and Technology Publications, Lda.

2 GLOVE-BASED SIGN LANGUAGE RECOGNITION

Through the use of multi-sensor fusion technology, the glove-based sign language recognition system precisely obtains the angle information, movement trajectory, and temporal information of the hand in order to accomplish sign language classification and recognition (Amin, 2022). The system has a high recognition rate, but the equipment is bulky and costly, which affects the ease of use and naturalness of human-computer exchange and makes it difficult to be used in real life.

2.1 Representative Works

In recent years, researchers have developed a new type of data glove, which adds contact sensors and can effectively utilize the large amount of contact information in Chinese sign language. For example, the novel data glove proposed by Zhang Yaxin et al. is a device designed for Chinese sign language recognition, which fully takes into account the characteristics of Chinese sign language and has the advantages of cheap price and high recognition accuracy. This glove solves the deficiencies of existing data gloves when used for Chinese sign language recognition by improving the application of sensors and contact sensors. In addition, the design of this glove takes into account the user's need for sign complexity, and therefore the sensors have been adapted accordingly. The main objective of this research is to improve the accuracy and efficiency of sign language recognition to meet the needs in practical applications. Yaxin Zhang's team experimentally verified the feasibility of this new data glove in virtual environment interaction tasks and demonstrated its effectiveness in the field of teleoperation. This suggests that the glove is not only suitable for sign language recognition, but can also be useful in other applications that require precise gesture control. In conclusion, the novel data glove proposed by them significantly improves the performance of Chinese sign language recognition by optimizing the sensor configuration and design, and at the same time is cost-effective and easy to operate, which gives it a wide range of potential applications in a variety of application scenarios. In terms of sensor design, this new data glove adds contact sensors on the basis of the original, cancelling the thumb-crossing sensors and the abduction sensing between the middle finger and the ring finger and the ring finger and the little finger for measuring the finger tensor angle, and adding contact sensors at the end of the finger, which can obtain a total of 20 information points. Moreover, the system has low cost and can recognize Chinese sign language better. As for the depth information, it was found that the “depth information” reflected by the bending angle of the elbow joint is also indispensable, so the three aspects were considered together to obtain a new type of wearable human posture sensor. Compared with the widely used q-type and p-type data gloves, this data glove adds a new type of contact sensor, effectively utilizes a large amount of contact information in Chinese sign language, and has the advantages of being cheaper, more suitable for the characteristics of Chinese sign language, and higher recognition accuracy. Finally, this data glove is applied to a new Chinese sign language recognition system, which can recognize Chinese sign language words more accurately by combining with the visual part (Zhang, 2001).

This design has several advantages, first of all this new data glove has added contact sensors that can effectively utilize the large amount of contact information in Chinese sign language. This makes it possible to more accurately capture the details and subtle changes of gestures during the recognition process. Moreover, compared with the existing CyberGlove model data glove, the new data glove has a cost advantage and is better suited to the characteristics of Chinese sign language. This means that it is not only less expensive, but also more compatible with Chinese sign language habits in practical applications. By incorporating a visual component, the data glove is able to recognize Chinese sign language words more accurately. This high-precision recognition capability is crucial to improving the overall performance of the sign language interpreting system. Importantly, the glove is able to acquire and analyze finger contact information, thereby reducing repetitive or unnecessary contact information and improving recognition efficiency and accuracy. The new data glove also has the ability to differentiate between left and right fingers, which is important for certain specific sign language gestures, as different combinations of fingers may be required for different gestures. The new data gloves utilize common and inexpensive components such as Bluetooth modules, gyroscopes and flexible sensors, making the total cost significantly lower than current data gloves with similar capabilities. This not only reduces the cost of the device, but also simplifies the algorithm complexity. Finally, the glove implements a real-time recognition and decoding system on smart terminals, ensuring fast data processing, which is very important for the user experience in real applications.

In order to achieve real-time sign language recognition and translation on lightweight edge devices and to offer deaf people with real-time communication and exchange services anytime, anywhere, Yin Yafeng's group has proposed another representative work that achieves this goal. The technique is based on area-aware time-sequence maps. This technology aims to realize real-time sign language recognition and translation on lightweight edge devices to provide anytime, anywhere communication and exchange services. Specifically, the technology utilizes computer vision and image processing techniques to acquire images of sign language movements through a cell phone camera or other video capture device. The system then uses algorithms to analyze and process these images to recognize finger and hand positions and movements. During the recognition process, the technique may incorporate a finite state machine and dynamic time regularization (DTW) approach to deal with continuous gesture movements in sign language videos. In addition, it may also involve deep learning models, such as the Keras deep model, for classifying and recognizing sign language actions. Ultimately, the recognized sign language movements are translated into text or speech and displayed via a connected digital screen or other output device so that they can be understood by the hearing impaired and others. In terms of hardware design, edge devices typically need to have high-performance computing power to support complex tasks, while requiring low power consumption and a small footprint. For example, the NVIDIA Jetson Xavier NX is a lightweight device pre-installed with Ubuntu, easy installation, and support for 12-24V wide-voltage operation and -10~55°C wide-temperature operation. These devices are also equipped with abundant input and output ports, which facilitate the connection of various types of sensors for multi-stream video edge inference and obstacle avoidance. For software architecture, lightweight implementation frameworks such as Cafe2, add support for mobile devices and mainstream machine learning frameworks such as PyTorch and MXNet are starting to be deployed on edge devices. EdgeOS, the IoT edge operating system, is built specifically to adapt to edge-side devices, with core functions such as industrial protocol parsing, data filtering and distribution, and is characterized by cross-platform, ease of use and secondary development. The last is to think about how to optimize the algorithm, researchers in the field of image recognition and other areas, through the design and optimization of lightweight convolutional neural network model, can be achieved on the edge device real-time image recognition tasks. Modifications like quantization, pruning, and knowledge distillation can help cut down on computing overhead and model size even further. A combination of a lightweight decoder and a pyramid pooling transformer for edge intelligence captures spatial and spectral details in the shallow layers via wavelet transforms to effectively recognize edges and reduce noise while maintaining computational efficiency. Another infrared weak target detection algorithm for embedded edge computing devices first uses a lightweight backbone network for feature extraction, and then obtains the final segmented bipartite map through multiple up-sampling layers and feature fusion across layers, which ensures high detection accuracy and low false alarm rate (Yin, 2017).

2.2 Discussions on Glove-Based Sign Language Recognition

Reducing the cost of data gloves while maintaining or enhancing their recognition accuracy can be achieved through several strategic approaches. One such approach is optimizing sensor configuration, which involves eliminating unnecessary sensors like the thumb-crossing sensor and the abduction sensor that measures the angle of finger spread. These sensors can be redundant in certain applications and contribute to increased costs. Additionally, replacing high-cost sensors with low-cost contact sensors can lead to significant cost reductions due to the reduced price of the sensors and the associated signal conversion circuitry.

Improving algorithms and data processing techniques is another vital strategy. The utilization of machine learning algorithms, such as the generalized regression neural network (GRNN), can significantly boost gesture recognition accuracy, with research indicating potential achievement of up to 99% accuracy. Furthermore, incorporating neural network models and template matching techniques can enhance the recognition rate of similar gesture letters, with the algorithm achieving a recognition rate of 98.5%.

The adoption of high-precision, low-latency sensors also play a crucial role in enhancing recognition precision. Selecting sensors that offer high accuracy and optimizing their arrangement on the glove can improve the stability and reliability of data capture. Advanced technologies like magnetic fingertip tracking sensors and electron magnetic field magnetic localization tracking can provide highly accurate finger motion capture data.

Lastly, simplifying the structural design of the data glove can contribute to cost reduction and maintenance ease. Designing the glove with a removable outer fabric allows for easy replacement of the outer layer in case of staining, reducing maintenance costs and prolonging the glove's lifespan. These combined efforts not only make data gloves more affordable but also ensure they remain effective tools for gesture recognition.

3 IMAGE-BASED SIGN LANGUAGE RECOGNITION

The vision-based sign language recognition system utilizes a camera to acquire 2D images or videos of sign language, and recognizes them through algorithms such as image processing and machine learning (Wiley, 2018). The system is closer to the social needs and suitable for human-computer interaction, but it has the defects of low recognition rate, poor real-time performance, and low applicability (Subburaj, 2022). In recent years, with the development of deep learning technology, the performance of vision-based sign language recognition system has been significantly improved.

3.1 Representative Works

Wuhan University proposes an attention-based mechanism for continuous sign language recognition algorithm attention-based 3D convolutional neural network (ACN). 3D CNN is mainly used to process data containing temporal dimension, such as video or 3D images. Unlike 2D CNNs, 3D CNNs use a 3D convolutional kernel that is capable of performing convolutional operations in the time dimension to capture spatio-temporal features in the data. For example, in video analysis, 3D CNNs can take into account both frame-to-frame motion information to better understand video content. By dynamically assigning varying weights to different areas of the input data as it is processed, the attention mechanism, on the other hand, allows the model to focus on the most relevant information, mimicking the functioning of the human visual system. The attention mechanism in deep learning can greatly enhance the model's performance, particularly when handling complicated and high-dimensional input. Finally, introducing the attention mechanism into 3D CNNs can improve the performance of the model by making it more focused on the critical parts of the input data. For example, in EEG signal emotion recognition, a 3D CNN that combines the frequency-space attention mechanism (FSA-3D-CNN) is able to simultaneously consider the information of EEG signals in the three dimensions of time, space, and frequency, thus improving the accuracy of emotion recognition. It is capable of recognizing continuous sign language in complex backgrounds. The algorithm preprocesses sign language videos containing complex backgrounds through a background removal module and extracts spatio-temporal fusion information using 3D-ResNet based on spatial attention mechanism (Yang, 2023).

The team of Prof. Hongwen Cao and Prof. Hong Li from the School of Foreign Languages, Chongqing University has made a new progress in the neural mechanism of Chinese sign language word recognition. This study examined the effects of word frequency, word length, phonological neighborhood word size, and likelihood on vocabulary recognition in sign language and found that these factors were similar to findings in spoken language, suggesting that the same neural mechanisms exist in the process of vocabulary recognition in sign language and spoken language. However, the significant effect of likelihood also suggests that the lexical recognition process is also influenced by factors related to linguistic modality. This study enriches the understanding of the neural mechanisms of sign languages in China, contributes to the further understanding of the nature of natural language, and provides important information about the characteristics and patterns of lexical processing in Chinese sign languages for both educators and learners of sign languages. Recent advances in sign language recognition technology include methods combining sequence annotation and deep learning, sign language recognition and translation techniques for region-aware time-series maps, continuous sign language recognition algorithms based on attentional mechanisms, and research on neural mechanisms (Zhang, 2023).

3.2 Discussions on Image-Based Sign Language Recognition

Improving the robustness of sign language recognition systems in complex environments is crucial for accurate interpretation. One approach to enhance robustness is through multimodal data fusion, which leverages a combination of multiple sensors and data sources. For instance, the integration of CNNs with inertial measurement units and stretchable strain sensors can more precisely perceive hand poses and motion trajectories. Utilizing a variety of multimodal data, including video feeds, keypoints, and optical flow, allows for the training of a unified visual backbone that significantly boosts recognition performance.

Deep learning model optimization is another key strategy. Advanced models such as the BLSTM (Bidirectional Long Short-Term Memory) model decompose consecutive sentences into word vectors, thereby enhancing the recognition of continuous sign language sentences. The fusion of attention mechanisms with connective temporal classification methods enables the extraction and combination of short-term spatio-temporal features and hand movement trajectory features. This addresses challenges related to redundant information and alignment issues within the spatio-temporal dimension.

To tackle the challenge of recognizing sign language from non-specific individuals, data enhancement and diversification are essential. Expanding the training dataset to include a broader range of signers improves the system's ability to generalize. Techniques like image generation for data augmentation can further strengthen the model's robustness, ensuring high accuracy in real-time recognition scenarios.

Lastly, the introduction of prior knowledge, including motor and linguistic a priori, into the causal temporal recognition framework is beneficial. This incorporation refines the robustness of feature extraction by providing a deeper understanding of the contextual semantics and the nuances of sign language gestures. By integrating these strategies, sign language recognition systems can be made more resilient and effective in complex and varied environments.

4 CONCLUSIONS

A significant application of computer vision and machine learning technologies in the field of accessible communication is the recognition of sign language using an image-based system. The performance of the sign language recognition system will continue to improve with ongoing technological advancements, providing the hearing impaired with a more convenient and effective means of communication. In the future, it is required to continue in-depth research to solve the current problems and promote the further development of sign language recognition technology. The system has a broad application prospect. It can not only provide more communication opportunities for the hearing impaired to help them communicate with others without barriers, but also can be applied in the field of education to help teachers and students understand sign language better and improve the teaching effect. In addition, the sign language image recognition system also has potential application value in the fields of intelligent transportation, remote control, and virtual reality.

REFERENCES

Amin, M. S., Rizvi, S. T. H., & Hossain, M. M. 2022. A comparative review on applications of different sensors for sign language recognition. *Journal of Imaging*, *8*(4), 98.

Cheok, M. J., Omar, Z., & Jaward, M. H. 2019. A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics*, *10*, 131-153.

Jones, G. A., Ni, D., & Wang, W. 2021. Nothing about us without us: Deaf education and sign language access in China. *Deafness & Education International*, *23*(3), 179-200.

Subburaj, S., & Murugavalli, S. 2022. Survey on sign language recognition in context of vision-based and deep learning. *Measurement: Sensors*, *23*, 100385.

Wadhawan, A., & Kumar, P. 2021. Sign language recognition systems: A decade systematic literature review. *Archives of computational methods in engineering*, *28*, 785-813.

Wiley, V., & Lucas, T. 2018. Computer vision and image processing: a paper review. *International Journal of Artificial Intelligence Research*, *2*(1), 29-36.

Yang, G., Ding, X., Gao, Y., et al. 2023. Continuous Sign Language Recognition with Complex Backgrounds Based on Attention Mechanism. *Journal of Wuhan University (Natural Science Edition), 69(1)*, 97-105

Yin, Y., 2017. Research on Behavior Perception Recognition Technology and System Based on Mobile Devices. Nanjing University. Doctoral Dissertation.

Zhang, Y., Yuan, K., & Yang, X., 2001, A Novel Data Glove for Sign Language Recognition. *Journal of University of Science and Technology Beijing, 23(4)*, 379-381

Zhang, X., Cao, H., & Li, H. 2023. Neurophysiological effects of frequency, length, phonological neighborhood density, and iconicity on sign recognition. *NeuroReport*, *34*(17), 817-824.