Research on gesture recognition technology based on machine learning

Yuang Xiong

Computer science, University College London, London, WC1E 6BT, United Kingdom

zcabioa@ucl.ac.uk

Abstract. Hand Gesture Recognition (HGR), as a significant technological advancement in the field of Human-Computer Interaction (HCI), aims to develop systems capable of accurately recognizing and interpreting human gestures for a diverse range of applications, including device control, virtual reality, gesture passwords, and gesture interaction. With the continuous advancement of machine learning algorithms, especially deep learning techniques, machine learning-based gesture recognition has garnered widespread attention. This paper presents a review of the development of gesture recognition techniques from traditional approaches to the current mainstream deep learning-based methods, and outlines the challenges and technical difficulties encountered. It analyzes several of the most popular classification techniques, including Naive Bayes, K-Nearest Neighbors (KNN), Random Forests, XGBoost, Support Vector Classifiers (SVCs), and Convolutional Neural Networks (CNNs). Furthermore, this paper examines the application of these algorithms in both dynamic and static gesture recognition and compares their performance and suitability in different scenarios. The results demonstrate that the accuracy and robustness of gesture recognition systems can be markedly enhanced through the prudent selection and optimization of the algorithms, which serves as a valuable reference for future research and applications.

Keywords: Machine Learning, Deep Learning, Hand Gesture Recognition (HGR), Human-Computer Interaction (HCI).

1. Introduction

Gesture recognition, as a natural and intuitive method of human-computer interaction, enhances direct communication by allowing gestures to be entered directly into the computer, thereby eliminating the need for intermediary devices. There are two main types of gesture recognition: static and dynamic. Static gesture recognition identifies predefined gestures, while dynamic gesture recognition conveys more nuanced information through the gesture movements [1]. The current methods employed for gesture recognition include vision-based recognition, data glove systems and depth-based recognition. Vision-based approaches often struggle under adverse conditions, and data gloves, although fast and accurate in capturing gesture information, are cumbersome and inconvenient to wear [2]. In contrast, depth camera-based recognition methods are gaining popularity due to their robustness, real-time performance, and high accuracy [3]. The widespread use of Kinect devices has significantly reduced the cost of depth-camera technologies, further fueling research interest [3]. Depth cameras, utilizing time-of-flight (TOF) and structured light technologies, can accurately capture three-dimensional object features, regardless of lighting conditions, shadows, or color variations [3]. Consequently, depth camera-based gesture recognition technology is expected to play a pivotal role in future research and applications [4]. This paper investigates deep learning-based gesture recognition techniques, focusing on the use of depth cameras and deep learning algorithms to enhance accuracy and robustness, particularly in complex environments. Additionally, the research aims to address current technological limitations, improve the accuracy and user experience of gesture recognition systems, and provide innovative solutions for future human-computer interaction technologies.

2. Overview of Gesture Recognition Technology

Gesture recognition refers to the process of identifying and understanding human intentions or messages conveyed through hand or body movements using technologies such as computer vision and machine learning [5]. By analyzing features such as the shape, motion, and speed of gestures, they can be translated into corresponding language or control commands, enabling interaction with computers or other smart devices [5]. Gesture recognition technology can be broadly categorized into static hand gesture recognition and dynamic hand gesture recognition. Static hand gesture recognition entails the identification of predefined static gestures, wherein the hand maintains a fixed position and shape, thereby enabling the user to control a computer or device through a specific gesture. In contrast, dynamic gesture recognition encompasses the recognition of gestures comprising a series of hand movements, the interpretation of which hinges on the trajectory and movement of the hand. This approach is more practical as it can discern continuous gestures in addition to static gestures, although it does present greater challenges. Related researches have developed a variety of widely used methods for gesture recognition, each of which has its own advantages and disadvantages, among which depth-based recognition methods are considered promising due to their high robustness, real-time performance and accuracy [5].

The design principles of gesture recognition technology involve several steps, including data collection, data pre-processing, feature extraction and optimization, classification algorithms, gesture recognition and control, and real-time and accuracy optimization. The collection of gesture data is typically accomplished through the use of sensors, such as data gloves, or cameras. Sensors on data gloves accurately convert finger and hand movements into real-time digital data, while cameras capture gesture shapes and movements through image processing techniques. In the pre-processing stage, the collected gesture data undergo noise filtering, normalization, and feature extraction to enhance the accuracy of gesture recognition algorithms. The process of feature extraction may entail the extraction of joint angles of fingers, hand contours, motion trajectories, and other relevant data points. The employment of enhanced particle swarm optimization (PSO) algorithms can facilitate the optimization of kernel parameters in support vector machines (SVM) through the utilization of pre-processed data. Subsequently, an appropriate classifier algorithm is selected, including Naïve Bayes, K-Nearest Neighbor (KNN), Random Forests, XGBoost, Support Vector Classifier (SVC), Logistic Regression, Stochastic Gradient Descent Classifier (SGDC), and Convolutional Neural Networks (CNN). The efficacy of different algorithms varies depending on the specific application scenario, necessitating the selection of an appropriate classifier that aligns with the desired requirements. The process of gesture recognition and control entails the classification of identified gestures through the use of trained classifiers, which can subsequently be employed to control devices or interact with computers. Ultimately, a variety of optimization techniques are utilized to augment the real-time functionality of gesture recognition systems. For example, the use of parallel computing helps to accelerate data processing, the implementation of efficient feature extraction algorithms reduces the computational load, and the continuous improvement of classifier models improves recognition accuracy [1, 5].

3. Applications of Machine Learning Algorithms in Gesture Recognition

Machine learning techniques are a prominent feature of gesture recognition, particularly in key processes such as feature extraction, classification, and recognition. Through machine learning algorithms, gesture recognition systems can learn gesture features from extensive data, thereby achieving efficient and accurate gesture recognition. The following will detail the applications of traditional machine learning methods and deep learning methods in gesture recognition.

3.1. Traditional Machine Learning Algorithms for Gesture Recognition

Traditional machine learning algorithms in gesture recognition are specified as follows.

Firstly, the Naïve Bayes algorithm, a simple and efficient probabilistic classification method based on Bayes' theorem and the assumption of feature independence, demonstrates robust performance and high computational efficiency when dealing with high-dimensional data, which is especially applicable to static hand gesture recognition. Secondly, the KNN algorithm is an instance-based learning method that calculates distances between new gesture data and samples in the training dataset to find the nearest K neighbors, using majority voting for classification. Despite its simplicity and intuitive nature, KNN incurs significant computational costs when handling large datasets. Thirdly, SVM is capable of identifying the optimal classification hyperplane for the differentiation of disparate categories of gesture data, exhibits robust generalization capabilities, and is well-suited for small sample learning. The recognition accuracy may be further enhanced by optimizing the SVM kernel parameters through the implementation of a modified particle swarm optimization (PSO) algorithm. A random forest is an ensemble learning method comprising multiple decision trees. The method summarizes the results by majority voting in order to achieve the final classification. This method exhibits high classification accuracy and robustness and is capable of effectively handling high-dimensional and complex data. Fourthly, XGBoost , an efficient gradient boosting tree algorithm that improves classifier performance through weighted voting, has performed well in many machine learning competitions with high computational efficiency and predictive accuracy. Fifthly, Logistic Regression is a generalized linear model used for binary classification problems, determining the best decision boundary by fitting data. While logistic regression performs well with linearly separable data, its effectiveness diminishes on complex nonlinear problems. Finally, SGDC optimizes model parameters for classification by iterative gradient descent and is suitable for large-scale datasets. However, it requires meticulous parameter selection and data pre-processing.

In conclusion, although traditional machine learning algorithms are widely used and exhibit satisfactory performance in the field of gesture recognition, it is of great importance to recognize that each method possesses distinctive advantages and limitations. The performance and applicability of gesture recognition systems can be further enhanced by combining the characteristics of different algorithms.

3.2. Deep Learning Algorithms for Gesture Recognition

In light of the accelerated advancement of deep learning technology, an increasing number of studies have commenced utilising deep learning methodologies with the objective of enhancing the efficacy of gesture recognition. Several commonly used deep learning algorithms and their applications in gesture recognition are listed below. Convolutional Neural Networks (CNN) are deep learning models specifically designed for processing image data. By using convolutional layers to extract local features of images and then using pooling layers and fully connected layers for classification, CNNs automatically extract multi-level features from gesture images, achieving high recognition accuracy. Long Short-Term Memory Networks (LSTM), a specific type of Recurrent Neural Network (RNN), is particularly adept at processing time series data. For dynamic gesture recognition, LSTM captures temporal features of gestures, accurately recognizing continuous gesture movements. Generative Adversarial Networks (GAN) consist of generators and discriminators that compete to generate high-quality data. In gesture recognition, GAN can be used for data augmentation to generate more training samples, thereby improving the generalization ability of the model. Transfer learning (TL) entails the process of adapting existing deep learning models to accommodate new gesture recognition tasks. It effectively utilizes knowledge from pre-trained models, thereby reducing the dependence on large amounts of annotated data and accelerating the training process. These deep learning algorithms applied in gesture recognition not only enhance recognition accuracy and robustness but also provide new avenues for further development and application of gesture recognition technology.

3.3. Practical Application Cases

At present, most hand gesture recognition techniques employ a combination of the above algorithms. For instance, Tam et al. proposed a real-time gesture recognition system that employs an embedded CNN for classification [6]. The integration of high-density surface electromyography (HD-EMG) and deep learning techniques enhances system reliability and reduces response time. The principal benefit of this approach is its simplicity, which is achieved by reducing the infrastructure necessary for classifier training. The method employs EMG to quantify the activity of pertinent upper limb muscles, obviating the necessity to train machine learning algorithms for gesture recognition. A comparison of this method with existing algorithms and techniques, including KNN, NB, DA, SVMs, and Random Forests (RFs), has demonstrated its utility in controlling virtual objects and wheelchairs. In addition, Lee and Tanaka elucidated the principles of gesture and fingertip recognition as they pertain to the development of natural user interfaces. Their discourse centered on the processes of recognition and tracking between gestures, wherein they highlighted the role of Kinect and depth sensing technologies in enabling gesture recognition even in environments with limited illumination or complex backgrounds. This, they posited, would facilitate the creation of natural communication interfaces [7]. Allard et al. employed deep learning for gesture recognition, underscoring its capacity to process a multitude of features within expansive datasets [8][9]. Despite the relatively limited use of deep learning algorithms in EMG-based gesture recognition, their study demonstrated that the collection of data from multiple users can effectively reduce recording errors and enhance the accuracy of gesture recognition. The researchers investigated the utility of extensive data sets for gesture learning and evaluated three distinct deep learning networks with inputs derived from raw EMG, spectrograms, and continuous wavelet transform (CWT) [10]. It was demonstrated that real-time feedback enables users to modify their muscle activation strategies, thereby attenuating the typical decline in accuracy that is typically observed during prolonged use [11].

4. Current Challenges and Future Prospects

4.1. Challenges for Gesture Recognition Technology

Based on the analysis of existing technologies, several major challenges are faced by gesture recognition systems. These systems typically demonstrate satisfactory performance in typical lighting conditions, but their accuracy is diminished in environments with complex backgrounds and obstacles. Modifications in environmental circumstances have a considerable impact on the resilience and dependability of the system, particularly in outdoor settings or environments characterized by significant fluctuations in illumination. Furthermore, the considerable discrepancies in muscle characteristics between amputees and individuals with intact extremities also contribute to the diminished accuracy of classification. In light of the discrepancies in hand morphology, dimensions, and kinematic patterns, it is imperative that the system exhibits augmented adaptability to accommodate the evolving gesture characteristics of diverse users.

The necessity for efficient real-time processing presents an additional challenge, which requires the optimization of algorithms to reduce the consumption of computational resources while maintaining the desired level of accuracy. This is of particular importance in the context of mobile or embedded systems, where the available computational resources and battery life are typically constrained. Moreover, it is of paramount importance to address the issue of accuracy drops during gesture transitions. This necessitates the development of enhanced transition handling mechanisms, such as gesture holding or time-series models, which aims to improve the user experience during continuous gesture input. And it is of paramount importance to guarantee the stability and accuracy of the system over an extended period of use. This is necessary to prevent a decline in performance over time, which could result from factors such as sensor wear, changes in user habits, and parameter drift. Finally, user privacy and data security should be protected throughout the collection and processing of hand gesture data, requiring the implementation of strong data protection policies to reduce the risk of data leakage and misuse.

4.2. Future Development Trends

In the future, hand gesture recognition technology will undergo significant advancements, particularly in a number of key areas. Firstly, it will be applied to practical interactive functions. such as gaming, Google Street View, and Google Earth, integrated with head-mounted displays (HMDs) to create 3D interfaces that provide a more natural user experience. The integration of gesture recognition with virtual reality (VR) and augmented reality (AR) technologies has the potential to enhance immersive interaction experiences. The potential of TL algorithms to enhance the performance of Convolutional Neural Networks (ConvNets), particularly in the context of out-of-sample gestures, will be examined. Further development and testing of the TL algorithm will facilitate the inclusion of upper limb amputees and address the challenges posed by muscle variability. The use of GANs for data augmentation will result in the generation of a more diverse set of training samples, thereby enhancing the model's ability to generalize. The investigation of transfer learning algorithms that utilize labeled information for long-term classification across different sessions will enhance the utility of the system. The continuous updating and adaptation of the system to user behavioral patterns will ensure the maintenance of efficient recognition performance. The adaptability of CNN models and computational platforms enables the system to accommodate diverse user capabilities and requirements, including varying types and quantities of gestures, which facilitates the development of a personalized prosthetic control solution for each user.

In addition, the use of adaptive deep learning models will be explored to dynamically adjust the recognition parameters according to changes in user habits. Improving the quality of HD-EMG datasets through cross-domain preprocessing methods will improve learning and classification results. Development of a fully embedded system will reduce data transfer and external computation requirements and improve real-time responsiveness. Future work will focus on developing efficient compression algorithms and edge computing techniques to reduce data processing latency and bandwidth requirements. The installation and usage process from sensor installation to classifier training will be simplified, minimizing user complexity, enhancing the intuitiveness and reliability of the system, and improving user execution and responsiveness in daily life. This will be achieved by simplifying the process from sensor installation to classifier training. The continuous optimization of the system interface and interaction modes based on user feedback will result in an overall improvement in user satisfaction. The integration of multiple sensor data types, including vision, depth, and EMG signals, will facilitate more precise and resilient gesture recognition. The integration of multiple sensor modalities will address the inherent limitations of individual sensors, thereby enhancing the overall performance of the system. The development of a hand gesture recognition system that is capable of automatically adapting to changes in the surrounding environment through real-time adjustments to the recognition algorithm using environment-aware technology will enhance the system's adaptability in a range of scenarios. Dynamic adjustment of system parameters using environmental sensors and machine learning techniques will further optimize performance in complex environments. By addressing these challenges and adapting to future trends, gesture recognition technology will be widely used in various fields to provide a more natural, efficient and reliable human-computer interaction experience.

5. Conclusion

In conclusion, the paper presents the definition and design principles of machine learning-based gesture recognition technology, and provides an overview of the most commonly used machine learning algorithms and their applications. The analysis of these algorithms and projects allows for an examination of the current challenges facing artificial intelligence technology and potential future directions. Hand gesture recognition technology, as a pivotal domain in the field of human-computer interaction, will persist as a focal point for future technological advances, with a major emphasis on machine learning-based gesture recognition techniques, which will help to achieve more accurate and faster recognition results while adapting to a variety of input formats. Despite demonstrating broad applications in fields such as healthcare, virtual reality, and smart homes, gesture recognition technology still faces challenges such as environmental variability, individual differences, real-time requirements, and long-term stability. Future development directions include the optimization of algorithms, the improvement of system adaptability and robustness, the enhancement of data processing and transmission methods, and the improvement of the user experience. The advancement of hand gesture recognition technology through continuous innovation will facilitate the development of intelligent and convenient systems, thereby enabling more intelligent and humanized forms of interaction.

References

[1] Chen, L., et al. (2013) A Survey on Hand Gesture Recognition. 2013 International Conference on Computer Sciences and Applications, 313-316.

[2] Moeslund, T.B., et al. (2001) A Survey of Computer Vision-Based Human Motion Capture. Computer Vision and Image Understanding, 81(3): 231-268.

[3] Bhushan, S., Alshehri, M., Keshta, I., Chakraverti, A.K., Rajpurohit, J. and Abugabah, A. (2022) An Experimental Analysis of Various Machine Learning Algorithms for Hand Gesture Recognition. Electronics, 11(6): 968.

[4] Oudah, M., Al-Naji, A. and Chahl, J. (2020) Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. J. Imaging, 6: 73.

[5] Tulli, S., et al. (2024). Hand Gesture Recognition: A Contemporary Overview of Techniques. 2024 International Conference on Automation and Computation (AUTOCOM), 457-463.

[6] Tam, S., Boukadoum, M., et al. (2019). A Fully Embedded Adaptive Real-Time Hand Gesture Classifier Leveraging HD-sEMG and Deep Learning. IEEE Transactions on Biomedical Circuits and Systems, 14(2): 232-243.

[7] Lee, U., and Jiro, T. (2013) Finger Identification and Hand Gesture Recognition Techniques for Natural User Interface. Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction, 274-279.

[8] Côté-Allard, U., Fall, C.L. Drouin, A., et al. (2019) Deep Learning for Electromyographic Hand Gesture Signal Classification using Transfer Learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(4): 760-771.

[9] Qi, J., et al. (2024) Computer Vision-based Hand Gesture Recognition for Human-robot Interaction: A Review. Complex Intell. Syst., 10:1581-1606.

[10] Senturk, Z.K, et al. (2021) Machine Learning Based Hand Gesture Recognition via EMG Data. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal.

[11] Qi, J, et al. (2004) Computer vision-based hand gesture recognition for human-robot interaction: a review. Complex & Intelligent Systems, 10(1): 1581-1606.