JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

HAND SIGN RECOGNITION

DHIVYA SOBIYA M, PADMASHREE T, ZAINAB HIRA J, PRABAKARAN G

STUDENT

VIVEKANANDHA COLLEGE OF TECHNOLOGY FOR WOMEN

ABSTRACT

The goal of the Hand Sign Recognition project is to create a reliable and effective system that can accurately read human hand motions. Through user-friendly gesture-based controls, this initiative seeks to improve human-computer interaction and bridge communication barriers for those who use sign language. In order to enhance model performance, the system first gathers a varied dataset of hand movements, after which preprocessing methods such frame extraction, hand detection, and data augmentation are applied. Convolutional neural networks (CNNs) and recurrent neural networks/long short-term memory (RNN/LSTM) networks are two examples of deep learning models that are taught to effectively recognize and categorize a variety of hand movements. In order to ensure accurate gesture interpretation, the system design also incorporates feature extraction techniques to record important hand characteristics. To enable smooth interaction and display recognized motions, an intuitive user interface is created. In order to make the solution inclusive and accessible to a broad spectrum of users, the project also places a strong emphasis on scalability, security, and privacy. Potential uses include virtual reality, gaming, smart home control, and sign language translation, demonstrating the adaptability and significance of the suggested technology.

Keywords: Hand Gesture Recognition, Deep Learning, CNN and RNN/LSTM, Human-Computer Interaction

1.1 HAND GESTURE RECOGNITION

In order to comprehend particular commands, actions, or communication patterns, hand gesture recognition technology records, analyzes, and interprets human hand movements. Usually, a mix of image capture, hand segmentation, feature extraction, and classification steps are involved. The system may link movements to established meanings, like sign language letters or symbolic acts, by recognizing static or dynamic hand postures. This technology creates opportunities in a number of areas, including instructional tools, contactless user interfaces, and assistive communication for the hard of hearing. Gesture recognition is an essential part of improving interactive systems because it can precisely identify and distinguish minute changes in hand position and motion.

FIGURE 1. HAND GESTURE RECOGNITION

1.2 DEEP LEARNING

By putting input through several processing layers, deep learning is a sophisticated family of machine learning algorithms that automatically finds complex structures in data. Deep learning models, which draw inspiration from the neural networks found in the human brain, can process vast amounts of intricate, high-dimensional data. Deep learning makes it easier to identify subtle hand shapes, textures, and movement patterns in hand gesture detection without requiring a lot of human feature creation. By learning hierarchical representations, models like CNNs and LSTMs are able to recognize long-term sequential patterns as well as fine-grained spatial details. Deep learning algorithms give increased accuracy, flexibility to new motions, and robustness against changes in background, lighting, or hand appearance as datasets expand and diversify.

1.3 CNN AND RNN/LSTM

By using convolutional operations to identify local patterns like edges, curves, and textures, Convolutional Neural Networks (CNNs) are experts in analyzing visual data. This makes CNNs particularly successful for analyzing static images of hand movements, extracting relevant spatial elements that distinguish one gesture from another. Conversely, Long Short-Term Memory (LSTM) networks and Recurrent Neural Networks (RNNs) are made to work with sequence data, identifying the connections between consecutive frames or hand positions. This ability is essential for identifying gestures that consist of a series of movements. In order to improve their classification performance across a variety of gesture sets, gesture recognition systems combine CNNs for feature extraction with RNNs/LSTMs for sequence modeling. This allows them to get a deeper knowledge of the structure and flow of hand gestures.

1.4 HUMAN-COMPUTER INTERACTION

The study and creation of systems that enhance how people interact with technology is the main goal of human-computer interaction, or HCI. Systems become more accessible, intuitive, and user-friendly by incorporating gesture detection into HCI, going beyond conventional tools like keyboards and mice. By allowing users to intuitively express their goals through movement, gesture-based HCI reduces obstacles to control and communication. It improves user experiences in settings where touch-based input is impractical and offers substantial benefits to those with physical disabilities. Additionally, by developing more fluid and adaptable interfaces that are suited to human behavior and preferences, gesture-driven HCI promotes more immersive applications in domains including gaming, education, healthcare, and virtual environments.

2. LITERATURE REVIEW

According to Danilo Avola [1] et al., in human relationships, hands are a powerful way to convey information that, in certain circumstances, can be used as a good substitute for speech, as is the case with Sign Language. Multimedia and computer vision researchers have long been interested in hand gesture detection. Sets of feature vectors that vary over time can be used to depict these gestures. Recurrent neural networks (RNNs) are ideal for analyzing these kinds of collections because of their ability to replicate long-term contextual information in temporal sequences. In this study, the angles formed by the human hand's finger bones are used to train an RNN. Because most human motions result in joint movements that produce incredibly unique corners, the desired characteristics—obtained via a Leap Motion Controller (LMC) sensor—were picked. The proposed approach and the effectiveness of the selected angles are evaluated on a challenging subset of many motions specified by American Sign Language (ASL). Additionally, the recommended method's superiority in hand gesture recognition accuracy has been demonstrated by comparison with previous state-of-the-art studies utilizing the SHREC dataset. In order to connect with innovative applications, including interactive games, hand gesture recognition offers a method for comprehending the data represented by the reported categories.

In this work, Junfu Pu [2] et al. have suggested A novel deep neural network and iterative optimization method for continuous real-world sign language recognition are described in this study. A visual input encoder for feature extraction and a sequence learning model that learns the relationship between the input sequence and the output sentence-level labels are the standard components of a continuous sign language recognition system. We use a 3D residual convolutional network (3D-ResNet) to extract visual information. The mapping between sequential properties and text phrases is then learned using a layered dilated convolutional network using Connectionist Temporal Classification (CTC). Since the CTC loss has little effect on the initial CNN parameters, training the deep network is challenging. We created an iterative optimization method to train our design in order to solve this problem. Using a sequence learning model, we create pseudo-labels for video clips and refine the 3D-ResNet while being monitored by pseudo labels to enhance feature representation.

Necati Cihan Camgoz [3] and colleagues proposed simultaneous alignment and recognition problems (often referred to as "sequence-to-sequence" learning) in this study. We deconstruct the problem into a series of specialized expert systems called SubUNets. The problem is subsequently solved by simulating the spatiotemporal interactions between these SubUNets, which are trainable all the way through. The method offers a number of significant advantages and is comparable to human learning and educational processes. SubUNets provide us to provide the system with domain-specific expert knowledge about suitable intermediate representations. Additionally, they allow us to access a wider range of more varied data sources by facilitating implicit transfer learning between many related activities. Our tests demonstrate that each of these attributes significantly enhances the overall recognition system's performance by more effectively limiting the learning problem. In the challenging subject of sign language recognition, the proposed solutions are shown. Our hand-shape recognition performance is state-of-the-art, surpassing previous methods by over 30%. Furthermore, without the need for an alignment step to separate the signs for recognition, we may attain sign recognition rates that are equivalent to those of previous studies.

In this paper, SALEH ALY [4] et al. have proposed. Because of its many applications in robotics, gaming, virtual reality, sign language, and human–computer interaction, hand gesture recognition has caught the attention of many researchers. The most efficient way for people with hearing impairments to communicate is through sign language, which is a systematic set of hand gestures. Hand segmentation, hand form feature representation, and gesture sequence recognition are the three main issues in developing an efficient sign language recognition system to differentiate dynamic isolated motions. A Hidden Markov Model (HMM) is used to identify sequences, hand-crafted feature extraction is used to describe hand forms, and color-based hand segmentation methods are used to segment hands in traditional sign language identification techniques. Based on a variety of deep learning architectures, such as hand semantic segmentation, hand form feature representation, and a deep recurrent neural network, this study suggests a novel framework for signer-independent sign language identification.

In this paper, Parul Wadhwa [5] et al. have suggested: The primary means of disseminating information is communication. Sign language is a means of communication for the dumb and deaf. Sign language is hard for a layperson to grasp. In order to help the deaf and dumb communicate with the rest of the world, we propose a hand glove that has flex and three-axis accelerometer sensors. The ARM LPC2148's template matching technique is used to compare the data collected by the sensors with predetermined values. A Wi-Fi module is then used to send the obtained data to an Android app. The alphabets used in American Sign Language are the main emphasis of this system. The vocal output for each character is played, and the data sent to the Android app is displayed on the screen. Humans are endowed with the capacity for speech. Some people, though, lack this talent. They only use sign language to communicate. Sign language is a communication method that uses a variety of hand and finger movements. To facilitate communication between the deaf and dumb and the rest of the world, several sign languages, including ASL and ISL, have been developed worldwide. In India, there are over 2.4 million people who are deaf or dumb.

3.EXISTING SYSTEM

There are numerous systems for using gestures to control robots. Certain gesture recognition systems employ morphological filtering, adaptive color segmentation, blocking to identify and label hands, template matching, and skeletonizing to identify gesture actions. The gesture inputs are not dynamic due to template matching. Some systems use a machine interface device to communicate with the robot in real time. There are other options that provide input to the program without requiring physical contact, like speech or hand movements, however many input-giving technologies require physical contact. There are two methods for gesture recognition. Vision-based (a) and glove-based (b). The glove-based method uses gloves or sensors to recognize hand motions. The glove-based method makes use of accelerometers, flex sensors, and other parts. Both static and dynamic gestures are possible. Hand positions are involved in static motions, and cameras are used to capture the image. The gathered images are subjected to segmentation for analysis. Although the background may contain multiple objects in addition to the hand region, the skin recognition method extracts the skin area from the input image. The RGB color model is used for the webcam-captured image.

4. PROPOSED SYSTEM

The main goal of the suggested system is to create a framework for hand gesture recognition that reliably records, interprets, and categorizes human hand movements. To guarantee that a broad range of hand positions and motions are covered, it starts with data collecting from video recordings, live feeds, and pre-existing datasets. To improve the quality of the input data, preprocessing methods such hand segmentation, background subtraction, and frame extraction are used. Key attributes such as finger placement and hand shape are isolated using sophisticated feature extraction techniques. To improve recognition accuracy, deep learning models are used, such as convolutional neural networks (CNNs) for learning spatial features and recurrent neural networks (RNNs) or long short-term memory (LSTM) networks for modeling sequences. To increase the stability and dependability of gesture interpretation, the system additionally incorporates post-processing techniques including confidence thresholding and prediction smoothing. Users may interact with the system naturally and effectively thanks to an intuitive user interface that displays recognized motions. The architecture ensures safe data handling and compatibility across various computing systems by being built for scalability, performance optimization, and user privacy.

A. DATA COLLECTION AND PREPROCESSING

Data about hand gestures can be gathered from a number of sources, such as publicly accessible databases, live camera feeds, and video recordings. A wide variety of hand gestures that represent various sign languages and typical gestures used in daily interactions must be captured. In preprocessing, individual images with hand motions are isolated by extracting frames from live camera feeds or video recordings. Hand regions within each frame are detected and localized using computer vision techniques including skin color segmentation and background subtraction. The dataset is made more diverse by applying picture augmentation techniques including rotation, scaling, and flipping to improve model generalization.

B. FEATURE EXTRACTION HAND REPRESENTATION

To separate the hand for additional examination, the identified hand regions are separated from the background. The segmented hand regions are then used to extract pertinent information, with an emphasis on motion trajectories, finger locations, and hand shape. Particularly when it comes to 3D gestures, feature extraction techniques like histogram of oriented gradients (HOG), keypoints, or depth information can be used to make sure the system records accurate and insightful representations of every motion.

C. MODEL TRAINING DEEP LEARNING MODELS

In order to learn spatial information from hand gesture photos, convolutional neural networks (CNNs) are used in deep learning models. Long short-term memory (LSTM) networks or recurrent neural networks (RNNs) are also used to record temporal dependencies in successive hand motions. To speed up training and improve performance, transfer learning is applied by using pre-trained models—like those trained on ImageNet—as feature extractors or by fine-tuning them on the hand gesture dataset. To evaluate the model, the dataset is divided into training, validation, and test sets as part of the training pipeline. Backpropagation and optimization techniques like stochastic gradient descent are used to reduce the loss function when training deep learning models. Cross-validation and grid search algorithms are used to optimize hyper parameters such as learning rate, batch size, and network design.

D. GESTURE RECOGNITION GESTURE CLASSIFICATION:

An inference pipeline is used for gesture classification, and trained models use camera feeds or video streams to complete classification tasks. Using softmax activation for multi-class classification tasks, hand gestures are categorized according to the learnt characteristics and model predictions. In post-processing, smoothing methods like temporal filtering are used to lessen jitter and noise in detected motions. To increase the robustness and dependability of gesture recognition, a confidence threshold is also defined to weed out predictions with low confidence.

E. USER INTERACTION AND FEEDBACK USER INTERFACE

A graphical user interface (GUI) or command-line interface (CLI) that offers an accessible display output is used to show users the identified movements. To help users, feedback mechanisms can display motions that have been recognized, provide textual descriptions, or provide voiced interpretations. Touchless interaction is made possible by interaction modes, which let users engage with the system without making physical contact. Additionally, multi-modal interaction is allowed, which combines touch inputs or voice commands with gesture detection to create a more accessible and improved user experience.

F. SYSTEM INTEGRATION AND DEPLOYMENT DEPLOYMENT PLATFORMS

The system is implemented as stand-alone desktop programs that work with Linux, macOS, and Windows. Furthermore, frameworks like Flask or Django are used to create web-based interfaces, guaranteeing cross-platform compatibility. For the system to function, a regular CPU and webcam are the very minimum of hardware requirements. It is advised to leverage GPUs or cloud-based computing resources for enhanced performance and increased accuracy.

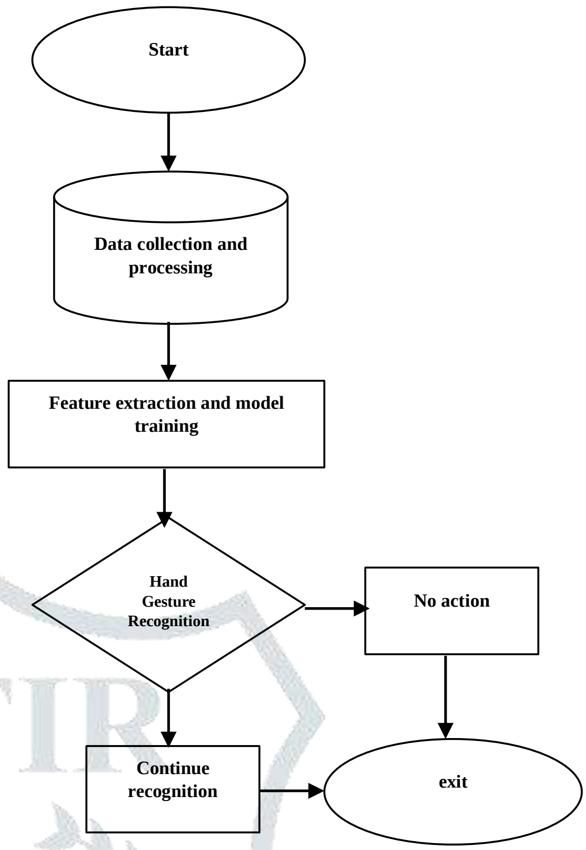

FIGURE 2. FLOW DIAGRAM

G. SCALABILITY AND PERFORMANCE OPTIMIZATION

To efficiently handle big datasets or heavy traffic loads, the system uses distributed computing techniques to spread across several nodes or cloud instances. By distributing incoming requests evenly among available resources, load balancing algorithms help to avoid bottlenecks. By using multi-core CPUs or GPUs to speed up compute processes like feature extraction or model inference, parallel processing approaches optimize performance. In order to minimize the size of deep learning models and maximize inference speed without compromising accuracy, model quantization approaches are also used.

H. SECURITY AND PRIVACY CONSIDERATIONS

The use of encryption methods to secure data transmission and storage ensures data security by shielding private information from unwanted access. In order to avoid data breaches and limit data access to authorized individuals, access control regulations are enforced. Anonymizing personally identifiable information from acquired data prioritizes user privacy and complies with data protection laws like the CCPA and GDPR. Transparency and control over the use of personal information are provided by the incorporation of permission mechanisms, which require users' express consent before collecting and processing their data.

5. ALGORITHM DETAILS

Data collection, which involves gathering pictures or videos of hand motions from multiple sources, is the first step in the Hand Sign Recognition algorithm. The next step is preprocessing, which involves removing frames from the movie and identifying hand regions using methods like skin color segmentation and background subtraction. Following detection, the hand regions are processed for feature extraction utilizing techniques like 3D depth information, keypoint detection, and histogram of oriented gradients (HOG). In order to learn spatial patterns, these features are fed into deep learning models such as Convolutional Neural Networks (CNNs). Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks are used to capture the temporal correlations between gestures. By optimizing previously learned models for the hand gesture dataset, transfer learning is used. To assess model performance, the data is separated into test, validation, and training sets. After training, the model uses the learnt features to classify hand gestures, and post-processing methods like confidence thresholding and gesture smoothing improve the recognition process. Through the use of parallel processing and distributed computing, the system is optimized for scalability. While the user interface offers feedback and visual representations of the identified movements, security features like data encryption and access restriction guarantee the protection of user data.

6. RESULT ANALYSIS

In order to analyze the Hand Sign Recognition system's results, a number of measures, including precision, recall, and F1-score, are used to assess the model's performance and accuracy. A different validation dataset is used to test the model's efficacy and make sure the system performs well when it comes to generalizing to invisible hand movements. The ability of several model architectures, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to accurately classify hand motions and manage complicated gestures is the main emphasis of the investigation. Furthermore, to improve overall robustness, post-processing methods such as gesture smoothing are used to reduce errors brought on by noisy data, and confidence thresholding is employed to weed out predictions with low confidence. Computational efficiency is another aspect of the analysis, where the system's processing time and resource consumption are assessed. To improve the system's scalability and speed, optimizations such parallel processing and model quantization are used.

7. CONCLUSION

To sum up, the Hand Sign Recognition system shows how deep learning models can effectively decipher hand gestures for communication. The system effectively and highly precisely classifies hand motions through meticulous data gathering, preprocessing, and feature extraction. The incorporation of sophisticated models such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) enhances the system's capacity to identify a variety of hand movements by enabling it to record both temporal and spatial characteristics. Performance and scalability are improved via optimization strategies including parallel processing and model quantization. Because of its design, the system can be implemented on several platforms while preserving user privacy and data security. In the end, this system offers a reliable and effective way to recognize hand signs, with room for improvement and possible use in a number of fields, including accessibility and human-computer interaction.

8. FUTURE WORK

There are numerous ways to improve the Hand Sign Recognition system in the future. Expanding the dataset to incorporate a greater range of hand motions, especially for less popular sign languages and different hand shapes, is one possible way to increase the model's accuracy. Furthermore, adding more sophisticated methods like 3D hand pose estimation could improve gesture detection even further, particularly in intricate situations with several hands or delicate motions. Experimenting with more complex deep learning architectures, including attention mechanisms, can also enhance the system's performance by better capturing significant hand motion information. Additionally, investigating the usage of more effective hardware accelerators, such as edge computing devices or specialized GPUs, may help the system scale and manage bigger datasets. The system's adaptability and accessibility could be improved in terms of user involvement by adding support for multi-modal inputs, such as fusing hand gestures with voice instructions or facial expressions. Last but not least, investigating integration with additional applications, including assistive technology for the hearing challenged or sign language translation systems, could increase the system's effect and practicality in everyday situations.

9. REFERENCES

1. D. Avola et al. identify sign language and semaphoric hand gestures using recurrent neural networks and a jump motion controller. 10.1109/TMM.2018.2856094, IEEE Trans. Multimedia, vol. 21, no. 1, pp. 234–245, January 2019.

2."Dilated convolutional network with iterative optimization for continuous sign language recognition," by J. Pu, W. Zhou, and H. Li, Proceedings of the 27th International Joint Conference on Artificial Intelligence, July 2018, pp. 885-891.

3. "SubUNets: End-to-end hand shape and continuous sign language recognition," by N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, in Proceedings of the IEEE Int. Conf. Comput. Vis. (ICCV), October 2017, pp. 3075–3084.

4. "DeepArSLR: A novel signer-independent deep learning framework for isolated Arabic sign language gestures detection," by S. and W. Aly, IEEE Access, vol. 8, pp. 83199–83212, 2020; doi:10.1109/ACCESS.2020.2990699.

5. "Sign Language Recognition using a Smart Hand Device with Sensor Combination," by Shrenika and M. M. Bala, in Proc. Int. Conf. Comput. Sci., Eng. Appl. (ICC SEA), March 2020, the first five pages.
6. "Movement epenthesis detection for continuous sign language recognition," by A. Choudhury, A. K. Talukdar, M. K. Bhuyan, and K. K. Sarma, in J. Intell. Syst., vol. 26, no. 3, pp. 471-481, July 5.

7. "Dynamic hand gesture recognition of Arabic sign language using deep convolutional neural networks," Proc. Conf. Signal Process. Commun. Eng. Syst. (SPACES), Jan. 2018, pp. 194-197, G. A. Rao, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry.

8. "A thorough review of hand gesture detection techniques, issues, and applications," by Yasen and Jusoh, PeerJ Comput. Sci., vol. 5, p. e218, Sep 2019.

9. "Review of constraints on vision-based gesture recognition for human-computer interaction," IET Comput. Vis., vol. 12, no. 1, pp. 3-15 Feb. 2018, doi: 10.1049/iet-cvi. 2017.0052, B. K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman.

10. "A Deep Learning Approach for Analyzing Video and Skeletal Features in Sign Language Recognition," Proc. Int. Conf. Imag. Syst. Techn., Oct. 2018, pp. 1-6, D. Konstantinidis, K. Dimittropoulos, and P. Daras.