International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Hand Gesture - Based Sign Language Interpreter

Kirubaharan . A

Department of Artificial Intelligence and data science, Bannari Amman Institute of technology

Sathyamangalam

kirubaharan.a@bitsathy.ac.in

***Abstract*—Sign language (SL) serves as the primary mode of communication for millions of individuals with hearing and speech impairments globally. However, it remains a language that is largely inaccessible to the general public, which poses a barrier to effective communication. This paper presents a static hand gesture-based sign language recognition system using Convolutional Neural Networks (CNNs). We utilize a dataset consisting of Indian Sign Language (ISL) gestures, including 26 alphabets and 10 digits. Our system uses image processing techniques for hand segmentation, followed by CNN-based classification. With a testing accuracy of 99.89% and validation accuracy of 99.85% over 5 epochs, our system proves to be effective and reliable in recognizing ISL gestures. This work provides a robust foundation for real-time applications aimed at facilitating communication between the hearing impaired and the general public.**

***Keywords*—–CNN, Sign Language Recognition, Indian Sign Language, Gesture Recognition, Deep Learning**

I. Introduction

Gesture recognition, an integral component of non-cognitive computing, allows computers to interpret human gestures. For hearing and speech-impaired individuals, gesture recognition, particularly through sign language, is crucial for communication. In India alone, around 63 million people suffer from hearing impairments, and the Indian Sign Language (ISL) is the primary mode of communication for a significant portion of these individuals. However, the complexity and diversity of ISL, which includes over 3000 signs, poses a challenge in making the language accessible to those unfamiliar with it.

This paper presents a solution using deep learning techniques to interpret static hand gestures. Leveraging Convolutional Neural Networks (CNNs), we propose a robust and accurate model capable of recognizing ISL gestures, including alphabets and digits. This work focuses on creating a practical, real-time system that can help bridge the communication gap between individuals with hearing impairments and the general population.

The paper is organized as follows:

Section II reviews related work,
Section III presents the proposed system,
Section IV discusses experimental results,
Section V concludes with future work.

II. Related Work

Several studies have explored hand gesture recognition and sign language recognition systems. However, only a few have employed deep learning techniques, which offer significant advancements in accuracy and efficiency.

Bhagat et al. [2] developed a system for ISL gesture recognition using CNNs. Their model employed five convolutional layers to achieve an accuracy of 98.81% on 36 static gestures. They also explored dynamic gestures using Convolutional LSTMs, achieving an accuracy of 99.08% on training video datasets.

Intwala et al. [3] created a dataset using MATLAB for ISL and applied pre-processing techniques followed by classification using the AlexNet architecture. They achieved a mean test accuracy of 87.69% for real-time images.

Sahoo et al. [4] proposed a system using Principal Component Analysis (PCA) for feature reduction combined with CNN for ISL gestures. Their model, using AlexNet for feature extraction, achieved a test accuracy of 99.32%.

In contrast to these works, our approach focuses on developing a CNN model trained with a larger dataset using Histogram BackProjection for image segmentation and classification. By capturing images of 26 ISL alphabets and 10 digits, we built a robust system that achieves near-perfect accuracy, as discussed in Section IV.

III. Proposed Work

In this section, we describe the methodology and design of our static hand gesture recognition system. The system consists of multiple phases, including hand segmentation, dataset creation, and CNN-based classification.

A. Hand Segmentation and Detection

One of the most critical steps in gesture recognition is segmenting the hand from the background. In this work, we used Histogram BackProjection for this purpose. Images are captured in real-time using a webcam, processed frame by frame, and then segmented. The steps for segmentation include converting the RGB image to HSV (Hue, Saturation, and Value), followed by morphological operations to remove noise and enhance the object of interest—in this case, the hand.

Figures 3a and 3b demonstrate the effectiveness of the segmentation process across different backgrounds. Our system showed optimal results when the user wore gloves, as seen in Figure 3d.

B. Dataset Creation

Given the lack of a comprehensive ISL dataset, we created our own dataset consisting of 26 alphabet gestures and 10 digit gestures. For each gesture, 1200 images were captured, resulting in a total of 2400 images per gesture after applying image augmentation techniques, such as flipping.

The final dataset comprised 72000 images of size 50x50 pixels. This dataset was then divided into training, testing, and validation sets, as shown in Figure 5. This division allowed for efficient training of the CNN model, ensuring accurate predictions during real-time testing.

C. CNN Architecture

We implemented a Sequential CNN model, as depicted in Figure 6. The CNN architecture includes multiple convolutional layers with Rectified Linear Unit (ReLU) activation functions, followed by MaxPooling layers. The final fully connected layer (FC) outputs the classification result.

The architecture of the model is optimized using the Stochastic Gradient Descent (SGD) optimizer with backpropagation. The combination of convolution, pooling, and dense layers enables the model to learn the spatial hierarchies in the input gestures, leading to high classification accuracy.

IV. Experimental Results

The system was trained using the dataset created in Section III. We experimented with different epochs to determine the optimal training configuration. Table I presents the accuracy achieved at various epochs, demonstrating that the highest accuracy of 99.89% was achieved at 5 epochs, as shown in Figure 9.

Epochs	Training Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)
3	91.55	99.33	98.95
5	94.01	99.89	99.88
7	94.36	99.60	99.36
9	95.60	98.92	99.15

From the results, we found that increasing the number of epochs beyond 5 led to overfitting, causing a slight decrease in accuracy. The computational time required for training also increased with the number of epochs, as shown in Table II.

Epochs	Time Taken (mins)
3	14.8
5	20.5
7	26.35
9	30.5

The system was also tested with live inputs, demonstrating real-time recognition of gestures, as shown in Figure 10. The confusion matrix, presented in Figure 11, shows the classification accuracy across all 36 classes.

V. DISCUSSION

The effectiveness of sign language recognition systems heavily depends on the accuracy of segmentation, feature extraction, and classification. The proposed system achieves notable improvements in each of these areas through the use of Histogram BackProjection for segmentation and Convolutional Neural Networks (CNNs) for classification. Our system recognizes static gestures with high accuracy, making it a viable tool for real-time communication with hearing and speech-impaired individuals. Below, we discuss key aspects of the system's performance and potential improvements.

A. Hand Segmentation

Hand segmentation is a crucial step in sign language recognition. In our system, the use of Histogram BackProjection provides a precise method for separating the hand from the background in real-time video streams. However, lighting conditions and the complexity of the background remain challenges. We observed that darker backgrounds and low-light environments affect the segmentation quality, leading to occasional misclassification. To address this, the system could incorporate adaptive background subtraction techniques or more advanced filtering algorithms that are robust against varying environmental conditions.

Moreover, while gloves provided improved accuracy in hand detection, they may not always be practical for users. Future enhancements could involve advanced skin color detection algorithms or the incorporation of depth sensors to improve hand segmentation without the need for gloves. Depth-aware cameras or stereo cameras could enhance the accuracy of segmentation, especially in dynamic environments with multiple objects.

B. Dataset Quality and Expansion

The dataset plays a pivotal role in the performance of any machine learning model. For this project, we created a custom dataset of 26 Indian Sign Language (ISL) alphabet gestures and 10 digits, resulting in 2400 images per gesture. The extensive dataset helped in achieving the high training accuracy observed in the CNN model. However, the dataset is limited to static gestures, which restricts the system's ability to recognize dynamic, sequential sign language gestures.

One significant advantage of our system is its flexibility in adding new gestures. This allows for the inclusion of additional words or phrases to improve communication. For example, the system could be expanded to include common phrases or dynamic gestures by capturing videos rather than static images, thus facilitating a more comprehensive sign language interpreter. Future iterations of this work could also involve collaborations with the ISL community to create a publicly available, large-scale dataset of both static and dynamic gestures.

C. CNN Architecture and Performance

The Convolutional Neural Network (CNN) employed in our system, with its multiple convolutional and pooling layers, efficiently extracts features from the input images, allowing for high classification accuracy. By using Rectified Linear Units (ReLU) as the activation function and Stochastic Gradient Descent (SGD) as the optimizer, the model is able to train quickly and effectively. Our model achieved a peak testing accuracy of 99.89% after 5 epochs, which is higher than many existing systems. The model’s performance, however, begins to decline after 7 epochs, likely due to overfitting.

One way to further improve the model’s generalizability and prevent overfitting is through techniques like dropout, where a fraction of the neurons in the fully connected layers are randomly ignored during training. Additionally, increasing the diversity of the dataset by introducing various lighting conditions, backgrounds, and hand orientations could help the CNN learn more generalized features, further improving its performance in real-world scenarios.

D. Real-Time Application and Practical Challenges

In real-time testing, the system performed well, with each letter and digit being correctly identified and displayed on the screen. However, several challenges emerged during practical implementation. First, the system's reliance on specific background conditions may limit its use in uncontrolled environments. Ensuring a light-colored, uniform background is not always feasible, especially in everyday settings.

Another practical consideration is the system’s reliance on static gestures. While useful for alphabet and digit recognition, real-world communication often involves dynamic gestures and sentences composed of multiple gestures. To overcome this limitation, future work could focus on integrating dynamic gesture recognition through the use of Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, which are well-suited for sequence prediction tasks.

Moreover, integrating speech synthesis into the system could provide an immediate auditory output, enhancing the overall communication process. This would allow hearing-impaired individuals to not only see their gestures translated into text but also have the text spoken aloud, further bridging the communication gap.

E. Comparison with Existing Systems

When compared with other gesture recognition systems, the proposed work demonstrates superior accuracy. As shown in Table III, existing systems achieve accuracy levels between 87.69% and 99.32%, whereas our system achieves an accuracy of 99.89%. This can be attributed to the careful design of the CNN architecture, the use of an extensive dataset, and effective segmentation techniques. However, most of these systems, like the one developed by Sahoo et al. [4], focus on American Sign Language (ASL) rather than ISL, indicating a gap in the availability of systems designed for Indian users.

Our system also stands out due to its flexibility in adding new gestures. By allowing users to expand the gesture database, we provide a tool that can evolve with the needs of the community, unlike some static systems that are limited to pre-defined datasets. This adaptability ensures that the system can remain relevant as new gestures and sign language standards emerge.

VI. Future Work

While the current system performs well for static gesture recognition, there are several avenues for future research and development. These include:

1) *Dynamic Gesture Recognition: Future iterations of this project will focus on the recognition of dynamic gestures. This would involve capturing sequences of images or videos and applying models like Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks to classify these sequences into words or phrases.*

2) *Improved Hand Segmentation: The current segmentation technique works well in controlled environments, but more robust segmentation methods are needed to handle varied backgrounds, lighting conditions, and occlusions. Integrating depth sensors or using more advanced computer vision techniques could significantly improve segmentation quality.*

3) *Multi-Modal Recognition: Integrating hand gestures with facial expressions and body movements could improve the recognition of more complex sign languages, which often involve more than just hand movements. This would require the use of additional sensors and more sophisticated models capable of interpreting multiple input streams.*

4) *Cross-Language Support: While this project focuses on Indian Sign Language, the system could be adapted to recognize other sign languages, such as American Sign Language (ASL) or British Sign Language (BSL), by retraining the CNN with corresponding datasets. This would make the system more versatile and applicable to a broader audience.*

5) *Portable and Wearable Solutions: To enhance usability, future versions of the system could be developed as a mobile application or embedded into wearable devices like smart gloves. This would allow users to communicate more easily on-the-go, without needing access to a desktop system.*

VII. Conclusion

This paper presented a CNN-based static hand gesture recognition system designed for Indian Sign Language (ISL). The system achieved a high testing accuracy of 99.89% with real-time recognition capabilities, demonstrating its potential as a practical communication tool for hearing and speech-impaired individuals. By using Histogram BackProjection for hand segmentation and a sequential CNN model for classification, the system effectively recognizes 26 alphabet gestures and 10 digits from ISL.

Despite its success, the system has limitations, particularly in handling dynamic gestures and varied backgrounds. Future work will focus on addressing these challenges by incorporating dynamic gesture recognition, improving segmentation techniques, and expanding the system’s capabilities to support multi-modal inputs and cross-language recognition.

Ultimately, this work represents a step forward in creating accessible, real-time sign language interpreters that can improve communication for millions of individuals worldwide.

References

[1] Bhagat, N.K., Vishnusai, Y., Rathna, G., "Indian sign language gesture recognition using image processing and deep learning," Digital Image Computing: Techniques and Applications (DICTA), IEEE, 2019.

[2] Intwala, N., Banerjee, A., Gala, N., "Indian sign language converter using convolutional neural networks," International Conference for Convergence in Technology (I2CT), IEEE, 2019.

[3] Sahoo, J.P., Ari, S., Patra, S.K., "Hand gesture recognition using PCA based deep CNN reduced features and SVM classifier," International Symposium on Smart Electronic Systems (iSES), IEEE, 2019.

[4] Sarkar, A., Talukdar, A.K., Sarma, K.K., "CNN-based real-time Indian sign language recognition system," Advances in Computational Intelligence and Informatics, Springer, 2019.

[5] Sruthi, C., Lijiya, A., "Signet: A deep learning-based Indian sign language recognition system," International Conference on Communication and Signal Processing (ICCSP), IEEE, 2019.