Sign Language Translator using Machine Learning

Rushikesh Khilare[1], Aaryan Chengaloor[2], Satyajit Pokale[3]
Under the Guidance of: Prof. Santosh Bhosale
Department of Computer Engineering, Prerna Pratishthan's Universal College of Engineering and Research, Sasewadi, Pune
Savitribai Phule Pune University, Pune
Emails: rpkhilare70@gmail.com, pokalesatyajit@gmail.com, aaryanchengaloor@gmail.com

Abstract

Sign language is one of the oldest and most natural forms of communication, yet most people do not understand it, making it difficult for deaf and mute individuals to communicate. This research proposes a vision-based machine learning model for real-time sign language translation using convolutional and recurrent neural networks. The system recognizes hand gestures representing the American Sign Language (ASL) alphabet with an achieved accuracy of 98%. The model employs image preprocessing, CNN-based feature extraction, and RNN-based temporal analysis to recognize dynamic gestures effectively. This paper highlights the implementation, results, and future improvements for improving accessibility and inclusivity.

Keywords — Sign Language, Machine Learning, CNN, RNN, ASL, Gesture Recognition.

Sign language is one of the oldest and most natural forms of communication, yet most people do not understand it, making it difficult for deaf and mute individuals to communicate. This research proposes a vision-based machine learning model for real-time sign language translation using convolutional and recurrent neural networks. The system recognizes hand gestures representing the American Sign Language (ASL) alphabet with an achieved accuracy of 98%. The model employs image preprocessing, CNN-based feature extraction, and RNN-based temporal analysis to recognize dynamic gestures effectively. This paper highlights the implementation, results, and future improvements for improving accessibility and inclusivity.

I. Introduction

Deaf and mute individuals rely on sign language for communication, which uses hand gestures, facial expressions, and body movements to convey meaning.

However, communication barriers persist as most people do not understand sign language. This project addresses this gap by developing a real-time sign language translator using machine learning techniques. Our model focuses on fingerspelling recognition using a CNN-RNN hybrid model for robust and accurate gesture classification.

American sign language is a predominant sign language Since the only disability Deaf and Dumb (hereby referred to as D&M) people have is communication related and since they cannot use spoken languages, the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behavior and visuals. D&M people make use of their hands to express different gestures to express their ideas with other people. Gestures are the non-verbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language. A sign language is a language which uses gestures instead of sound to convey meaning combining hand-shapes, orientation and movement of the hands, arms or body, facial expressions and lip-patterns. Contrary to popular belief, sign language is not international. These vary from region to region.

Sign language is a visual language and consists of 3 major components

Sign Language (SL) is definitely a boon to deaf and mute people for communicating in daily life. Using any SL like American Sign Language (ASL), Arabic Sign Language (ArSL) or any other, a person can convey messages by the help of movement of the hands rather than sound patterns. SL involves simultaneously defined shapes, orientation and movement of the body parts. The main problem is that the majority of healthy people have little to no understanding of sign languages. Thus, effective communication is considered a challenge and an obstacle for the deaf and mute in their daily lives. This proposed work is an attempt to overcome the communication barrier and discomfort in the society by providing the deaf people with a method of communication so that they can share their thoughts and feelings effectively and independently without the need for a translator. If on one side a deaf person communicates using the SL that he is familiar and comfortable with, and then the system translates this SL into sound and visuals, which the able person can understand, this would be an effective solution bridging the gap between the two different means of communication

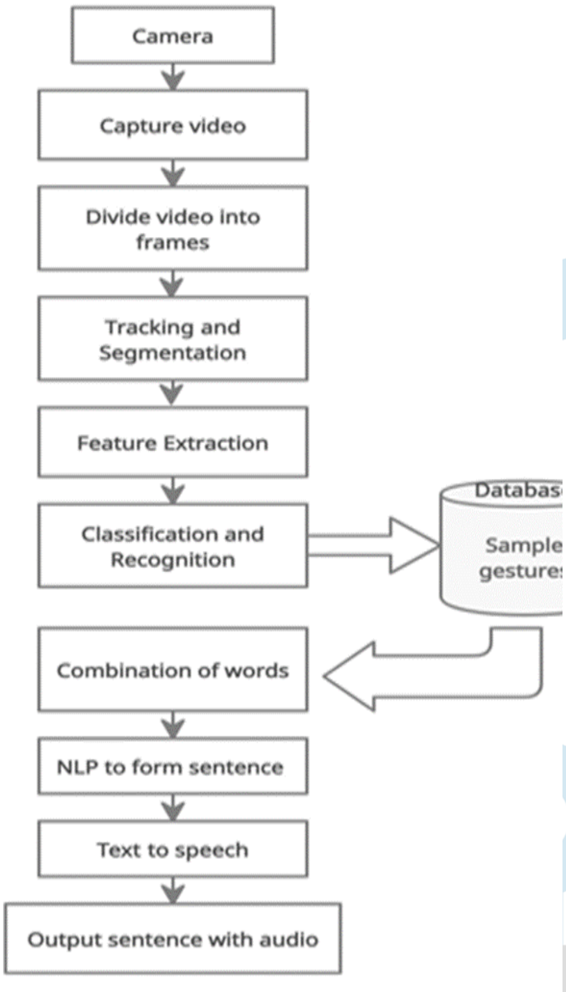

Figure 1. System Architecture of Real Time Sign Language Interpreter

II. Literature Review

Recent studies demonstrate multiple approaches to gesture recognition. Vision-based methods rely on image processing and neural networks, while sensor-based methods depend on hardware devices like gloves or Kinect. Research using CNNs and HMMs achieved recognition rates above 90% for static gestures. However, sensor-based approaches are costly. Our study builds upon vision-based CNN architectures for affordable real-time translation using a webcam.

Recognizing a sign language gesture from continuous gestures is a very challenging research issue. The researchers solved this problem using gradient based key frame extraction method. These key frames were helpful in splitting continuous sign language gestures into sequence of signs as well as for removing uninformative frames. After splitting of gestures each sign has been treated as an isolated gesture. Then features of pre-processed gestures were extracted using Orientation Histogram (OH) with Principal Component Analysis (PCA) is applied for reducing dimension of features obtained after OH. Experiments were performed on their own continuous ISL dataset which was created using canon EOS camera in Robotics and Artificial Intelligence laboratory (IIIT-A).

III. Methodology

The proposed model includes preprocessing, training, and classification phases:
1. Frame Extraction: Frames are extracted from input gesture videos.
2. Preprocessing: Background noise is removed using Gaussian Blur and thresholding.
3. CNN Model: Extracts spatial features using convolution and pooling layers.
4. RNN Model: Analyzes temporal features to identify gesture sequences.
5. Prediction: Combines outputs for accurate letter and word prediction.

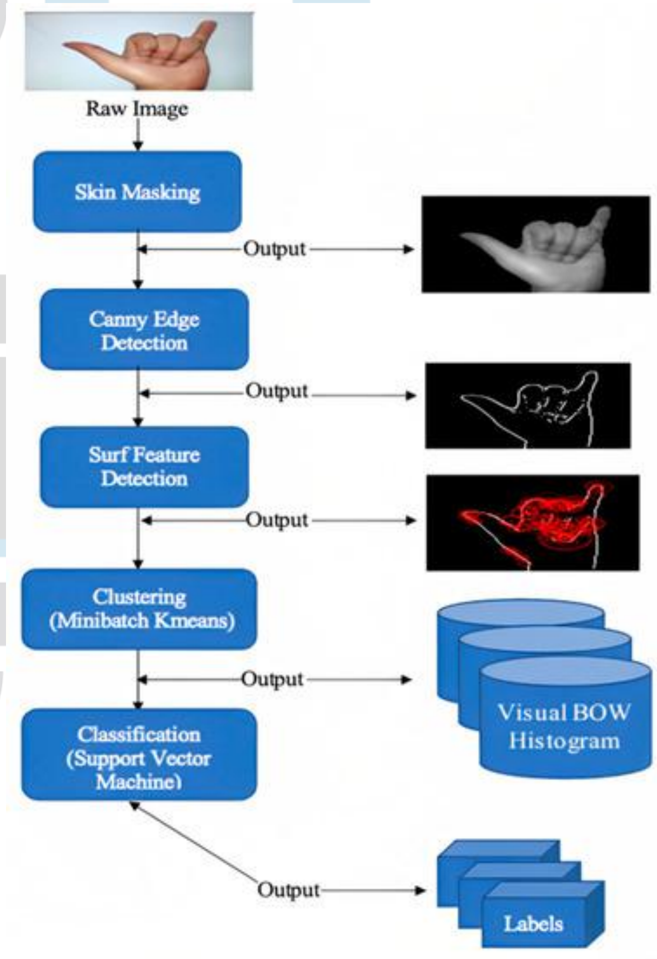

Figure 1. Flowchart of gesture-to-text translation.

IV. Dataset Description

A custom dataset was created using OpenCV, containing over 800 training and 200 testing images per ASL alphabet symbol. Images were captured via a webcam and processed into 128x128 grayscale frames. Gaussian Blur filtering enhanced hand-feature clarity before feeding into the CNN model.

For the project we tried to find already made datasets but we couldn’t find datasets in the form of raw images that matched our requirements. All we could find were the datasets in the form of RGB values. Hence, we decided to create our own data set. Steps we followed to create our data set are as follows.

V. Results and Discussion

The CNN-only model achieved 95.8% accuracy, while combining CNN and RNN improved accuracy to 98%. Comparative analysis shows our model outperforms other vision-based systems without requiring expensive hardware. The approach demonstrates strong real-time performance with minimal lag.

The training accuracy achieved when training the image dataset without any augmentation was very high (around 90 percent), but the real-time performance was not up to par. Most of the time, it predicted incorrectly because hand-gestures were not precisely centred and vertically aligned in real time. To compensate for this shortcoming, we trained our model by supplementing our dataset. Although the training accuracy was reduced to 89 percent, the real-time predictions were mostly correct. Offline testing of approximately 9000 augmented images revealed 82 percent accuracy.

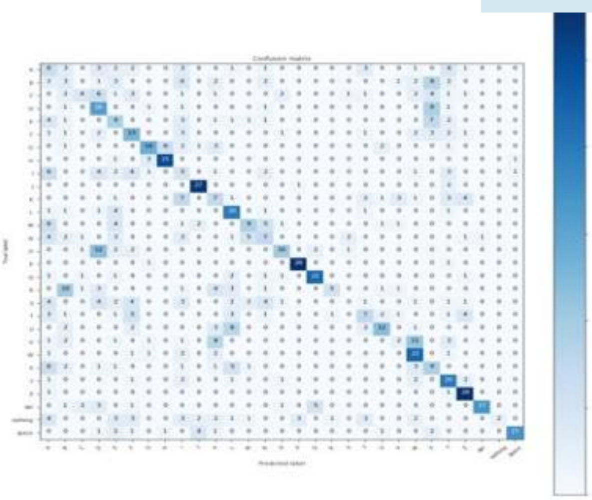

Fig -6: Confusion matrix.

The generated heatmap is based on the confusion matrix of the model, illustrating the accuracy of predictions for each sample.

VI. Conclusion

This research presents a machine learning-based sign language translator achieving 98% accuracy in recognizing ASL gestures. The integration of CNN and RNN enhances spatial-temporal recognition. The system facilitates communication for the hearing and speech impaired and can be expanded for other languages and continuous gesture recognition.

The In this report, a functional real time vision based American Sign Language recognition for D&M people have been developed for asl alphabets.

We achieved final accuracy of 98.0% on our data set. We have improved our prediction after implementing two layers of algorithms wherein we have verified and predicted symbols which are more similar to each other.

This gives us the ability to detect almost all the symbols provided that they are shown properly, there is no noise in the background and lighting is adequate.

Hand gestures are a powerful way for human communication, with lots of potential applications in the area of human computer interaction. Vision-based hand gesture recognition techniques have many proven advantages compared with traditional devices. However, hand gesture recognition is a difficult problem and the current work is only a small contribution towards achieving the results needed in the field of sign language gesture recognition. This report presented a vision-based system able to interpret isolated hand gestures from the Argentinian Sign Language (LSA).

VII. Future Scope

Future work will include improving detection accuracy under low-light and complex backgrounds, adding natural language processing (NLP) for contextual understanding, and developing mobile and web-based deployment for accessibility.

We are planning to achieve higher accuracy even in case of complex backgrounds by trying out various background subtraction algorithms.

We are also thinking of improving the Pre Processing to predict gestures in low light conditions with a higher accuracy.

This project can be enhanced by being built as a web/mobile application for the users to conveniently access the project. Also, the existing project only works for ASL; it can be extended to work for other native sign languages with the right amount of data set and training. This project implements a finger spelling translator; however, sign languages are also spoken in a contextual basis where each gesture could represent an object, or verb. So, identifying this kind of a contextual signing would require a higher degree of processing and natural language processing (NLP).

The potential for sign language translation using Convolutional Neural Networks (CNNs) is extensive and promising. Researchers can explore various avenues to enhance this technology, such as developing more extensive sign language datasets, refining real-time translation capabilities, integrating multi-modal data, accommodating regional sign language variations, and expanding to encompass different sign languages.

VIII. References

1. Pigou L. et al. (2015). Sign Language Recognition Using CNNs. ECCV 2014 Workshops.

2. Zaki M.M., Shaheen S.I. (2011). Vision-based Features for Sign Language Recognition. PRL.

3. Kang B., Tripathi S., Nguyen T.Q. (2015). Real-time Sign Language Fingerspelling Recognition using CNNs.

4. Mukai N., Harada N. (2017). Japanese Fingerspelling Recognition. IEEE NicoInt.

[5] http://www-i6.informatik.rwth-aachen.de/~dreuw/database.php

[6] Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham

[7] Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision-based features. Pattern Recognition Letters 32(4), 572–577 (2011).

[8] N. Mukai, N. Harada and Y. Chang, "Japanese Fingerspelling Recognition Based on Classification Tree and Machine Learning," 2017 Nicograph International (NicoInt), Kyoto, Japan, 2017, pp. 19-24. doi:10.1109/NICOInt.2017.9

[9] Byeongkeun Kang, Subarna Tripathi, Truong Q. Nguyen" Real-time sign language fingerspelling recognition using convolutional neural networks from depth map" 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)

[10] Number System Recognition (https://github.com/chasinginfinity/number-sign-recognition)

[11] https://opencv.org/

[12] https://en.wikipedia.org/wiki/TensorFlow