



Sign Language to Text Conversion

¹Prof.Deveshree Wankhede, ²Ashish Anil Jaiswal, ³Om Rajendra Nandiskar, ⁴Siddhant SunilShettigar

¹Assistant Professor, ^{2,3,4}Student^{1,2,3,4}Department of Computer Engineering, ^{1,2,3,4}Shivajirao S. Jondhale College of Engineering, Dombivli, India

Abstract : Hand gesture is one of the methods used in sign language for non-verbal communication. It is most commonly used by deaf & dumb people who have hearing or speech problems to communicate among themselves or with normal people. The proposed system would be a real time system wherein live sign gestures would be processed using image processing. Then classifiers would be used to differentiate various signs and the translated output would be displaying text. This project is designed to develop a machine learning model that can classify multiple identifiers used to fingerprint languages. In this model, the classification machine learning algorithm is trained on image data and tested on the data. The purpose of the system will be to develop a system prototype that will be able to automatically recognize sign language to help deaf and dumb people to communicate more effectively with each other or normal people.

Keywords: Preprocessing of Dataset, Post-gaussian effect, Gaussian Blur Filter, Convolutional Neural Network (CNN) model.

I INTRODUCTION

Computer Technology has tremendously grown over the past decade and has become a necessary part of everyday live. Most laptops today are equipped with webcams, which have recently been used in insecure applications utilizing face recognition. In order to harness the full potential of a webcam, it can be used for vision based. The usefulness of a webcam can also be greatly extended to other HCI applications such as a sign language database or motion controller.

Sign Language is the means of communication among the deaf and mute community. Sign Language emerges and evolves naturally within the hearing impaired community. There is no standardized sign language for all deaf people across the world. However, sign languages are not universal, as with spoken languages, these differ from region to region. A person who can talk and hear properly (normal person) cannot communicate with deaf & dumb person unless he/she is familiar with sign language.

Communication is a form of communication in which people with disabilities use their hands as different gestures to express their ideas to others. Gestures are nonverbal communication that can be understood visually. Non-verbal communication of hearing impaired people is called sign language.

Sign language is a visual language consisting of 3 main elements

Fingerspelling	Word level sign Vocabulary	Non-manual Features
Used to spell words letter by letter	Used for the majority of Communication	Facial expression and tounge,mouth and body position

Fig 1.1 Sign Language Components

Above table shows how our model will work and process In this project, I focused on creating a model that can recognize gestures based on fingerspelling to give each gesture a complete word.

The explanations I learned are shown below.

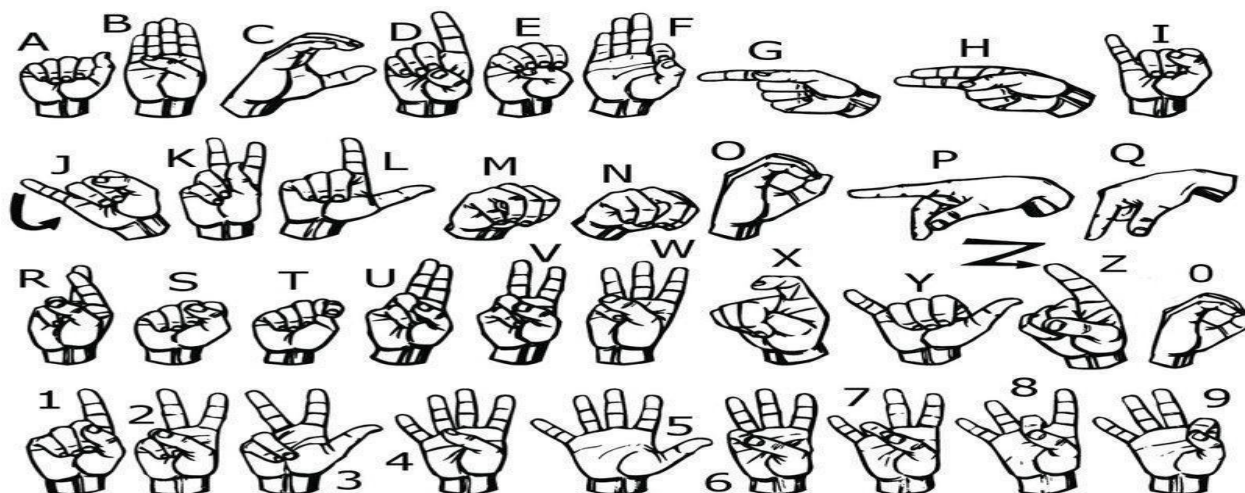


Fig 1.2 American Sign Language

American Sign Language (ASL) is a visual-gestural language used by the Deaf and hard of hearing community in the United States, employing handshapes, facial expressions, and body movements to convey meaning.[7]

II LITERATURE SURVEY

The First Paper “Application of Tensor Train Decomposition in S2VT Model for Sign Language Recognition”. The study introduced tensor-train decomposition to reduce the parameters in S2VT models for sign language recognition. It explored how different configurations impact performance. Experimentation found that applying tensor-train to the fully-connected layer and the first LSTM layer maintained high accuracy and significantly reduced parameters. However, representing the second LSTM layer with tensor-train decreased accuracy. Overall, tensor-train models significantly reduced parameters, beneficial for mobile deployment, but careful selection of layers for decomposition is crucial for maintaining accuracy. [1]

The Second Paper “EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and Artificial Intelligence,” The research paper presents EasyTalk, an application designed to translate Sri Lankan Sign Language into text and audio formats, as well as translate English text into Sri Lankan Sign Language. It comprises components for hand gesture detection, image classification, text and audio generation, and text to sign conversion, aiming to bridge the communication gap between the hearing-impaired and the general public. The system demonstrates high accuracy in detecting hand signs, classifying them, and generating corresponding text/audio or sign language outputs [2].

The Third Paper “Real-time Conversion of Sign Language to Text and Speech,” The paper discusses the development of an Android application that enables real-time conversion of American Sign Language (ASL) gestures into text and speech without the need for special gloves or sensors. The process involves utilizing vision-based techniques, particularly the use of image processing and machine learning, specifically focusing on Histogram of Gradients (HOG) descriptors and Support Vector Machine (SVM) for training the model [3].

The Fourth Paper “Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space,” the paper presents a novel approach in Continuous Sign Language Recognition (CSLR) by integrating text information with video data to enhance intra-gloss dependencies. By employing cross-modal alignment between video and text embeddings, the method improves CSLR by creating more descriptive video-based latent representations, leading to more accurate recognition. Experimental evaluations on major CSLR datasets demonstrate the approach’s efficacy, surpassing prior state-of-the-art methods [4].

The Fifth Paper “Real time Motion-Based Hand Gesture Recognition from Time-of-Flight Video, Springer Science and Business Media New York”, This paper proposes an innovative solution for real-time hand gesture recognition using Time-Of-Flight (TOF) video technology. The system detects motion-based gestures by modeling patterns based on human arm anatomy. Synthetic motion patterns are generated and compared to captured input patterns for classification. The evaluation includes a significant collection of gestures, providing 3D pattern classification results and comparing them with 2D information. The paper discusses potential applications, such as interaction with maps and multimedia menus. While the approach is innovative, the paper could benefit from more contextualization, usability evaluation, and detailed technical explanations.

The Sixth Paper “A Review on Indian Sign Language Recognition, International journal of computer applications, Vol. 73, pp: 22,” this paper focuses on Automatic Sign Language Recognition (ASLR), specifically targeting Indian Sign Language (ISL). Acknowledging the complexity of ISL, the researchers propose a three-step methodology: preprocessing, feature extraction, and classification. Noteworthy is the emphasis on double-handed signs in ISL, in contrast to single-handed sign languages like American Sign Language. The study explores classification methods like Artificial Neural Networks, Support Vector Machines, and Hidden Markov Models. While the merits include relevance and localization, the demerits encompass potential dataset biases, oversimplification, and the absence of ethical and comparative discussions.

III METHODOLOGY

Images for each alphabet will be captured 1199 times and will be processed and saved in the dataset after being resized, converted to grayscale, and applying the post-Gaussian effect during the processing stage. The System will then be Trained over these processed images to recognize and predict the sign. In this stage, the processed images from the processing stage will be provided as a dataset to the system to train itself over the available Data and Output. The system will then go into the testing phase where it will predict the unknown data or the data which is not provided in the dataset. It will attempt to predict the output based on the learning from the

previous outcomes.

CNN-Convolutional Neural Networks (CNNs) revolutionize pattern recognition by mimicking human visual processing. They use special techniques such as layers, overlays and overlays to remove features from the image. Filters in convolutional layers detect edges, textures, and shapes, while pooling layers downsample feature maps for computational efficiency and translational invariance. All layers classify the image based on the learning experience. CNNs transcend image recognition, impacting fields from healthcare to autonomous vehicles. Despite challenges like overfitting, ongoing research addresses limitations, propelling CNNs towards greater sophistication and applicability. CNNs exemplify the transformative potential of artificial intelligence, reshaping how machines perceive and interpret visual data.

IV EXISTING SYSTEM

Introduction

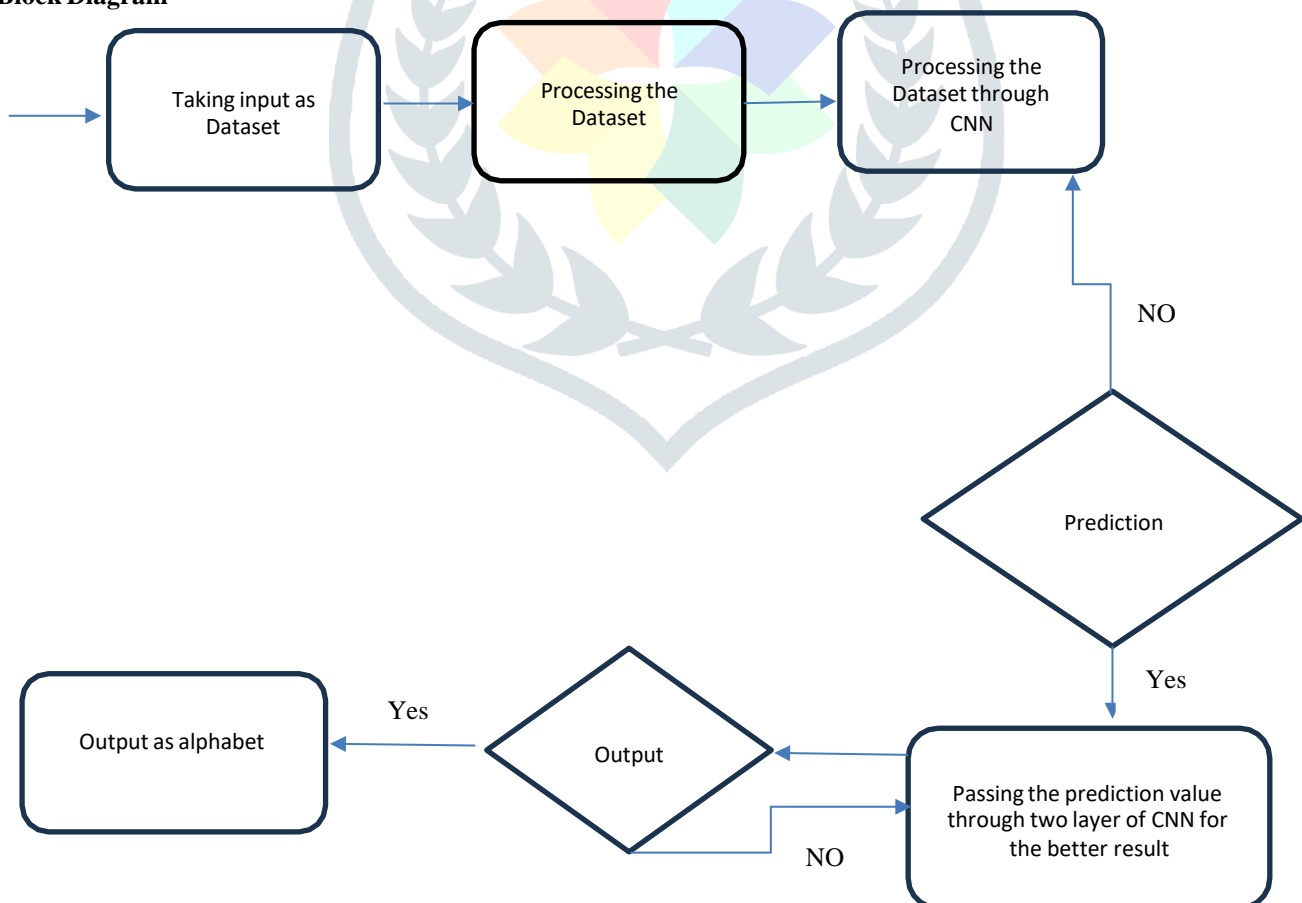
In the existing system, it is mainly concerned with American,Britain Sign Languages. It focuses on only 1 language. In other applications if a person wants to communicate with a normal person, then by using American Standard Sign Language they can communicate. By using this Application a normal person can understand the Sign language used by special people, our application will display Sign Language in text format which a normal person can read or listen to by clicking on the speak button. [1]

Focus Area

Several papers and projects have targeted the issue of hand gesture recognition.the main focuses of this Application is that it is very useful for Special people like Dumb and Deaf people. By using this application special people will gain importance as much as a normal person in every field focus on-

- Secured communication.
- Ensure data accuracy.
- Minimize manual data entry.
- Minimum time needed for the various processing.
- Greater efficiency.
- Better service.
- User friendliness and interaction.
- Minimum time required.

Block Diagram



Input as Dataset: This step involves collecting the dataset, likely containing images of sign language gestures.Preprocessing the Dataset: Data preprocessing is performed to clean, normalize, and prepare the dataset for training.Preprocessing the Dataset Through CNN: The preprocessed dataset is fed into a CNN, which learns features and patterns from the sign language

images.Prediction (loop) no: This indicates a one-time prediction rather than continuous or looping predictions.Passing the predict value through two layers of CNN for better results: The prediction from the initial CNN is passed through additional layers of a CNN for further processing and improved accuracy.Output (loop) no: Similar to step 4, this suggests a single output rather than a looping process.Output as Alphabet: The final output of the process translates the predicted sign language gesture into an alphabet or text, converting the sign language into readable text.

V LIMITATIONS OF EXISTING SYSTEM

- In the existing system,it is mainly concerned with American, Britain Sign Languages.
- It focuses on only 1 language.
- Color of background may be recognized as sign by computer, provided if green screen/black screen is not present as background
- As the majority of the people use their country's sign language to communicate with each other.

VI PROBLEM STATEMENT AND OBJECTIVES

Problem Statement

Communication is one of the basic requirements for survival in society. Deaf people use language to communicate, but their words are difficult for hearing people to understand. Many studies have been done on the recognition of American Sign Language, but Indian Sign Language is very different from American Sign Language. ISL uses both hands to communicate (20 out of 26 hands), while ASL uses one hand to communicate. Our goal is to take simple steps to bridge the communication gap between hearing and deaf people using Hindi Language. (Block_6) Good expansion of a word and expression not only allows deaf-mutes to communicate with the outside world more quickly and easily, but also contributes to the development of the path of self-understanding.

Objective

The aim of our project is to present an efficient and accurate mechanism to transcribe text or speech , thus the “dialog communication” between the deaf and hearing person will be smooth.

Sign Language is the means of communication among the deaf and mute community. Sign Language emerges and evolves naturally within the hearing impaired community. There is no standardized sign language for all deaf people across the world. However, sign languages are not universal, as with spoken languages, these differ from region to region. A person who can talk and hear properly (normal person) cannot communicate with deaf & dumb person unless he/she is familiar with sign language.

VII PROPOSED SYSTEM

Introduction

Most gesture recognition methods usually contain three major stages. The first stage is the processing model. In this model we will first convert the image from the dataset in grayscale image and the grayscale image is resized according to the requirement. Then the image is given a post-gaussian effect. Then the image is reshaped and saved in an array (X- features, Y- Label) is saved in x.pixel and y.pixel.

The Second Stage is the Training model. We opened the pickle file that we saved in process.py and we loaded them in 2D convolution layer and applied max pulling & flattened then saved it into a folder.

The Third Stage is the Testing Model We resize the image and applied the post-gaussian effect then with the help of the previously trained model we try to recognize the image and predict the output

- Most gesture recognition methods usually contain three major stages. The first stage is the processing model. In this model we will first convert the image from the dataset in grayscale image and the grayscale image is resized according to the requirement. Then the image is given a post-gaussian effect. Then the image is reshaped and saved in an array (X- features, Y- Label) is saved in x.pickle and y.pickle.
- The Second Stage is the Training model. We opened the pickle file that we saved in process.py and we loaded them in 2D convolution layer and applied max pulling & flattened then saved it into a folder.
- The Third Stage is the Testing Model We resize the image and applied the post-gaussian effect then with the help of the previously trained model we try to recognize the image and predict the output.

Algorithm/Process design

Algorithm-1

1. Pre-processing:

a. Gaussian Blur Filter:

- Gaussian blur filter is used by OpenCV to capture frames. This step helps in reducing noise and details in the image, making it smoother and easier for subsequent processing steps.

b. Thresholding:

- Thresholding is used to convert a grayscale image into a binary image where pixels are classified as black or white based on a property.This simplifies the image and highlights important features.

2. Feature Extraction:

- The processed image from step 1 is now ready for feature extraction. In the context of sign language, features may include the contours, shapes, and patterns formed by the hand gestures.

3. CNN Model for Prediction:

- The features extracted from the processed image are inputted into a Convolutional Neural Network (CNN) model. The CNN is trained to recognize patterns associated with different sign language letters or gestures.

- The model predicts the letter corresponding to the hand gesture in the frame.
4. Letter Detection Over Time:
- To ensure accurate predictions, a temporal aspect is introduced. If a letter is consistently detected for more than 50 frames, it is considered a valid detection. This helps in filtering out false positives and adds a level of confidence to the predictions.
5. Word Formation:
- Once a letter is detected and confirmed over a certain duration, it is printed and considered for forming a word.
 - Space between words is considered using a blank symbol. If there is a pause in the detection of letters for a certain duration, a space is added to denote the separation between words.

Overall Workflow:

1. Capture Frame: Use OpenCV to capture frames.
2. Pre-process Image: Apply Gaussian Blur and Thresholding.
3. Feature Extraction: Extract features from the pre-processed image.
4. CNN Prediction: Use a CNN model to predict the letter based on features.
5. Temporal Analysis: Track the letter detection over multiple frames.
6. Word Formation: Print and form words based on consistent letter detections.
7. Space Recognition: Use a blank symbol to denote spaces between words.

This overall process allows for the conversion of sign language gestures into text by analyzing the temporal patterns and consistently detected letters.

Algorithm-2

1. D vs. R and U:

- The issue here may be related to the similarity in the visual features of the symbols D, R, and U. These symbols may share certain characteristics, such as curved or straight lines.
- To improve discrimination, you may consider:
 - Enhancing the feature extraction process to capture more distinctive features.
 - Adjusting the parameters of your detection algorithm to be more sensitive to specific features that differentiate D from R and U.
 - Training the classifier with more diverse examples of D, R, and U, possibly with variations in hand gestures or angles.

2. U vs. D and R:

- Similar to the first case, U, D, and R may share common visual elements, leading to misclassification.
- Solutions:
 - Fine-tune the feature extraction process to focus on unique characteristics.
 - Optimize the classifier for better discrimination by providing more training samples and considering different perspectives of the symbols.

3. I vs. T, D, K, and L:

- The symbols I, T, D, K, and L might have overlapping visual patterns, causing misclassification.
- Potential solutions:
 - Refine feature extraction to emphasize distinct features of each symbol.
 - Augment training data with variations in hand positions and angles for each symbol.
 - Consider using a more advanced classifier or adjusting parameters to improve discrimination.

4. S vs. M and N:

- The confusion between S, M, and N suggests visual similarities that may be challenging for the classifier.
- Improvement strategies:
 - Revisit feature extraction and focus on aspects that differentiate S, M, and N.
 - Increase the diversity of training samples to cover various hand gestures and orientations.

VIII RESULT



Fig. 8.1 Training and Testing Outputs

Above image show the capturing row image, gray scale image and image post gaussian for training and testing the dataset.

IX CONCLUSION

Gestures are a powerful form of human communication and have many applications in human-computer interaction. Vision-based gesture recognition technology has many proven advantages over traditional devices. But gesture recognition is a complex problem, and the current work is only a small part of what needs to be done to achieve the desired results in the Confirm language. This demonstrates the visionary ability to translate the American Language Guidelines and convert them into text or speech. It can be seen that the received classification model is able to recognize all training movements, which is also an important part of the user freedom required for such systems. Carefully selected features combined with machine learning algorithms have proven to be so effective that it is possible to use them for all in-flight learning. As a future work, it is aimed to regularly improve the system and conduct experiments using the data processing message. The end result is that, although there is still a lot of work to be done in this area, the solutions have created a solid foundation for the development of a familiar user interface with a view-based guarantee. Sign language patterns can be easily changed and the system can be configured to train new expressions.

References

- [1] B. Xu, S. Huang and Z. Ye, "Application of Tensor Train Decomposition in S2VT Model for Sign Language Recognition," in *IEEE Access*, vol. 9, pp. 35646-35653, 2021, doi: 10.1109/ACCESS.2021.3059660.
- [2] D. Manoj Kumar, K. Bavanraj, S. Thavananthan, G. M. A. S. Bastiansz, S. M. B. Harshanath and J. Alosious, "EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and Artificial Intelligence," 2020 2nd International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 2020, pp. 506-511, doi: 10.1109/ICAC51239.2020.9357154.
- [3] K. Tiku, J. Maloo, A. Ramesh and I. R., "Real-time Conversion of Sign Language to Text and Speech," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 346-351, doi: 10.1109/ICIRCA48905.2020.9182877.
- [4] I. Papastratis, K. Dimitropoulos, D. Konstantinidis and P. Daras, "Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space," in *IEEE Access*, vol. 8, pp. 91170-91180, 2020, doi: 10.1109/ACCESS.2020.2993650.
- [5] Javier Molina, Jose Antonio Pajuelo and Jose N Martinez, *Real time Motion-Based Hand Gesture Recognition from Time-of-Flight Video*, Springer Science and Business Media New York 2015.
- [6] Anuja V.Nair, Bindu.V, A Review on Indian Sign Language Recognition, *International journal of computer applications*, Vol. 73, pp: 22, (2013).
- [7] https://www.researchgate.net/figure/Fingerspelling-in-American-Sign-Language-which-represents-26-letters-and-10-digits-with_fig1_346023992.