STM JOURNALS

International Journal of
Optical Innovations and Research

Volume 1, Issue 1, 2023
January–June
DOI (Journal): 10.37591/IJOIR

https://journals.stmjournals.com/ijoir

Research IJOIR

A Study of ROI Based Sign Language to Text Translation in Real Time Using Deep Learning

Soujanya S.[1]*, Ruhaima[2], Rajashree Nambiar[2], Sakshi[2], Shanthika[2]

Abstract

Since sign language is their primary form of communication, it plays a significant role in the lives of hearingand speech disabled people. However, since not everyone is conversant in sign language, it is challenging for the disabled tointeract with others daily. The sign language is made up of a variety of hand gestures that may stand in for a wide range of words and sentiments. The purpose of this study is to develop atrustworthy communication interpretation. Using image processing and Machine learning, this paper translates Indian Sign Language into legible output. So, we convert the sign languages into letters and then to words and to sentences which can be understood easily by the people who do not understand sign languages. So, this translation may be used for everyday communication and can be used to teach different gestures to autonomoussystems that use gestures as input.

Keywords: Gesture recognition, hand gesture scan, sign language, a KINECT sensor, MATLAB

INTRODUCTION

Sign Language is the most used to language used by deaf-dumb people to express their thoughts. The non- hearing-impaired people do not try to learn this sign language. This leads to miscommunication between deaf- dumb and normal people. We can reduce the difference between the deaf community if we program a system for the purpose of translating sign language into text format. The most prevalent and natural sign language used worldwide is American Sign Language (ASL). ASL uses hands to express signs. One of the main advantages of American Sign Language is less complexity and wider usage.

If the person does not have the knowledge of sign language, then the communication with impaired hearing person becomes difficult. As a result, it would be beneficial to develop a system which could generate text based on American Sign Language. To recognize 26 gestures from the American Sign Language, we designed a real-time system that could recognize 26 gestures using a hand gesture recognition system in order to generate texts using the gestures. The signs are captured by using mobile camera.

The World Federation of the Deaf estimates that, out of 7.7 billion people in the world, approximately 70 million people are deaf [1]. SL is like any other language, which is widely used for communicating with people with hearing.

loss. Approximately 300 SLs (Sign Language) are used by the deaf community all over the world. ASL is a gesture-based language. For communicating

***Author for Correspondence**
Soujanya S.
E-mail: soujusalvankar@gmail.com

[1]Assistant Professor, Department of Electronics and Communication Engineering, Shri Madhwa Vadiraja Institute of Technology and Management, Banta KalUdupi, Karnataka, India
[2]Student, Department of Electronics and Communication Engineering, Shri Madhwa Vadiraja Institute of Technology and Management, Banta KalUdupi, Karnataka, India

Received Date: July 28, 2023
Accepted Date: August 12, 2023
Published Date: September 17, 2023

Citation: Soujanya S., Ruhaima, Rajashree Nambiar, Sakshi, Shanthika. A Study of ROI Based Sign Language to Text Translation in Real Time Using Deep Learning. International Journal of Optical Innovations & Research. 2023; 1(1): 8–14p.

with people across the globe, one must learn the native SL because there is no international standard for sign language [2]. The interpretation of SL varies from region to region. There is an American Sign Language Syntax and Grammar that is like that of any other natural sign language that is used.

In section II will see the literature review and section III we have discussed the methodology. Section IV of the report discusses the findings, and Section V concludes.

LITERATURE REVIEW

Humans communicate with one another in order to share their thoughts, feelings, and experiences with others around them. But this is not possible in the case for deaf and mute people. Through sign language, communication is possible for a deaf and mute person. Such people use their hands to communicate with other people and also to convey their feelings to others. Existing projects has certain drawbacks such as they cannot identify number signs and they are just machine learning models.

In paper [3] to convert ASL signed hand gestures into text as well as speech using unsupervised feature learning toeliminate communication barrier with the hearing impaired and as well provide teaching aid for sign language. The Features from Accelerated Segment Test (FAST) and Speeded UP Robust Features (SURF) algorithms are the foundation of their system, which formulates the relationship between picture segmentation and object recognition. This system goes through several stages, including data collecting using a KINECT sensor, picture segmentation, feature identification and extraction from ROI, and supervised and unsupervised classification of images with K-Nearest Neighbor (KNN) techniques and conversion of text to voice (TTS). ASL photos totaling 500 were gathered as a training set. From the collection of images,a database of forty-nine (49) different signs was used. The introduced system achieved a 92% accuracy of supervised feature learning and 78% of unsupervised feature learning.

Indian sign language hand motions are recognized using MATLAB, which is used to do the sign language recognition in paper [4]. The four processes in the proposed system are pre-processing, hand segmentation, feature extraction, sign recognition, and sign to text. Currently, research efforts have primarily concentrated on identifying static indicators of ISL from captured pictures or video sequences.

The study [5] is based on the development and implementation of an application that transforms American Sign Language to text and allows it to be used everywhere and at any time. Image is captured by the camera and skin segmentation is done using YCbCr systems. To identify the sign, features are taken from the picture using HOG and classification is performed. To classify data, Support Vector Machine (SVM) is employed. The accuracy obtained by thissystem is 89.54%.

The paper [6] aims to design an application which will translate sign language to English in the form of text and audio. This application will acquire the image data using the webcam of the computer, it is then pre-processed using an algorithm and template matching is used for recognition. 6000 images of English alphabets are used in the database of the system. The system produces 88% accuracy approximately. Future research will include working towards implementing this model on a mobile application.

The publication [7] describes the creation of a sign language learning aid for beginners that uses hand detection. The approach is built on the explicit skin-color space thresholding method of skin-color modelling. The hand will be separated from the background by this predefined skin-color range. These photos were used as input for the Convolutional Neural Network (CNN) model, which classifies images. Keras was used for training of images. This approach is done in real-time and is used for fast computation.

The paper [8] identifies Indian sign language as an essential communication medium for communicating with deaf people. In this work, they proposea deep learning-based methodology for building an Indian sign language static alphabet recognition system. Here, they examine different currently used techniques for decoding sign language and put into practice a CNN architecture for decoding the ISL static alphabet from the binary silhouette of the signer's hand area.

According to paper [9], the suggested system would recognize Indian Sign language, translate it into voice, and then translate the speech into text in English and Malayalam, which will then be shown on a smartphone. The proposed system is portable and consume very low power and also provide improved accuracy. It converts Sign language with the help of Flex sensor, microcontroller and display it on Android phone using Bluetooth module HC 05. This project can be further developed to converts more words phrases. More effectively detecting and sensing the dynamic motions. This work can be extended to converts sign language to text and speech in other Indian Language like Hindi, Tamil using Android Studio.

A novel model for the challenge of dynamic sign language recognition based on 3D Residual Convent and Bi-directional LSTM networks was put out in article [10]. By analyzing the video sequence, the proposed model could effectively recognize different hand gestures by extracting video spatiotemporal features and analyzing features sequence. In the experimental result on DEVISIGN_D dataset and

SLR_dataset, different sign language could be accurately andeffectively distinguished. In future work, the sign sentence recognition for real world video in complex environment willbe considered for further research, and the evaluation of experiment should be considered with more characteristics, like robustness, sensitivity and so on.

In paper [11] the author presents an analysis of the performance of various techniques that have been used for the translation of sign language to text/speech format. The best approach is adopted following analysis. Based on this study, an Android app was created that can translate ASL (American Sign Language) signals into text and speech in real time. 27 classes have been in their model each corresponding to letters in the English language. The SVM model which is a supervised learning technique will classify the images into the 27 classes to yield a result. The newly developed technology has a 97% accuracy rate.

A sign-language translation glove is created in paper [12] utilizing a deep learning algorithm and auxetic-interlaced AIYS array. They categorized 2,600 sign language movements encompassing all 26 letters using the ANN algorithm, and their accuracy rate was 99.8% across the board. The smart glove with the AIYS-array installed is used to translate sign language. The moveable joints in the fingers and wrist of a knitted glove are stitched with a 16-element yarn sensor array to create this glove [13].

The small number of datasets is the key problem. In order to record a person's signing motions in both Indian sign language and American Sign Language (ASL) and convert them into appropriate text and voice in real time, we are also building an application to go along with the dataset. Also, improper gesture is leading to poor recognition of the sign.

PROPOSED METHODOLOGY

In this paper we have proposed a system which includes gestures that converts them into alphabets/text. We are using a gesture recognition model using deep learning method. This procedure involves application of different filters for better feature extraction. This algorithm is used to classify an image based on various characteristics and make it possible to differentiate it from its respective classes. This proposed system in Figure 1 is a vision-based approach which means that all the signs are represented with bare hands and therefore it eliminates the use of any other sensory devices to represent the signs.

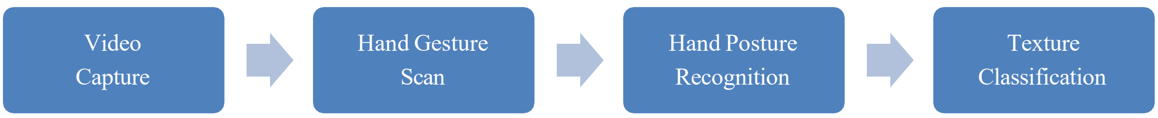

Figure 1. Block diagram.

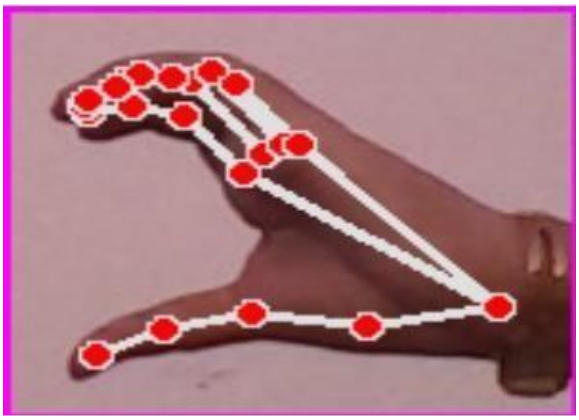

Figure 2. Dataset collected for letter “C” in ASL.

First, as seen in Figure 2, we took between 300 and 400 pictures of each of the ASL symbols. Then, using our computer's webcam, we record each frame that is displayed. An area of interest (ROI) is established for each frame and is represented by a square with a colored border. We take our ROI, which is RGB, from the entire image and turn it into a grayscale image. In order to extract different characteristics from our image, we then apply our Gaussian blur filter on it.

Next, the picture will be sent to CNN for prediction, following which the picture will be processed. There are spaces between each word in the form of the blank sign, which is used to indicate these spaces.

1. *First Convolution Layer:* A dimension of 128×128 pixel characterizes the input picture. In the primary convolutional layer, it undergoes initial processing using 32 filter parameters (3 ×3 pixels each). An image with 126×126 pixels will come out, one for each of those filter-weights. *First Pooling Layer:* We maintain the highest value in the 2×2 square of the array and down sample the images using maximum pooling of 2×2. As a result, our image has been downscaled to 63×63 pixels.
2. *Second Convolution Layer:* The output of the first pooling layer, which consists of these 63 x 63 pixels, is now used as the input for the second convolution layer.32 filter weights (3x3 pixels each) are used to process it in the secondconvolutional layer. A 60 by 60-pixel picture will be produced consequently.
3. The second pooling layer reduces the output pictures toa resolution of 30 x 30 with a maximum pool size of 2x2.
4. *First Densely Connected Layer:* There are now 28800 different combinations of values from the output of the second convolutional layer, and we use this array of values as an input to the layer of 128 neurons that are completely connected to the output of the second convolutional layer. The second densely connected layer receives the output of these layers. To prevent overfitting, we are utilizing a dropout layer with a value of 0.5.
5. *2nd Densely linked Layer:* A layer with 96 neurons that is fully linked to the output of the 1st Densely Connected Layer has been created to send the output from the top layer.
6. *The last layer:* The second densely connected layer's output feeds into the final layer, which has as many neurons as classes being classified (alphabets plus a blank sign).

RESULTS AND DISCUSSIONS

In this paper, we basically focus on producing fingerspelling-based hand gestures with number signs to form a complete sentence by combining each gesture for disabled people. To facilitate communication between mute people and others, the project is designed to use the predicted values in the android application for the mute person, to take advantage of their predicted values. The silent person will get access to the forecasts on their mobile device, allowing them to assess their correctness. As a result of these evaluations, future predictions can be further improved by using the information received as training input. In general, this approach narrowed the distance between silent people and others, enabling them to express emotions more and participate in society in a manner that they had not been able to do before.

We use a CNN classification system to categorize the 26 different forms of human hand motions to recognize them. A particular kind of feed-forward neural network is CNN. To allow the translation of the picture characteristics, small overlapping areas of the original image are represented by neurons in a layer of the network. Additionally, the convolutional layer's pixels use the same weight to reduce memory usage and boost effectiveness. CNN offers a rigorous and organized way to categorize the many types of hand motions, in contrast to earlier technologies that used challenging visual feature extraction.

The motions are captured by using a mobile camera that is mounted on a mobile device as presented in Figure 3. The frames from the stream are retrieved and converted to gray-scale images as shown in Figures 4 and 5. Hand motions are scanned which undergoes pre-processing before the model isfed for prediction. Data sets are provided to get the image detected. The gestures are translated into words.

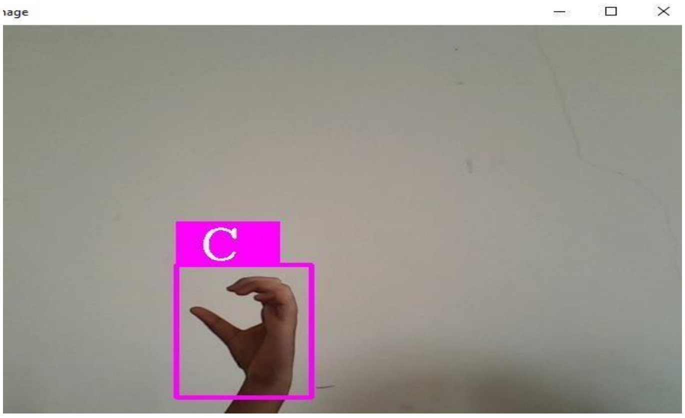

Figure 3. Recognized alphabet.

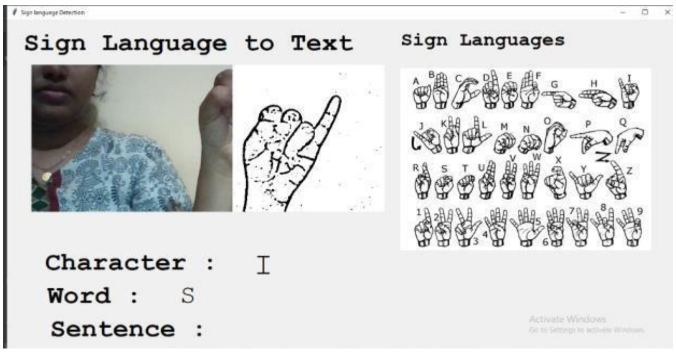

Figure 4. Sign language of letter ‘I’.

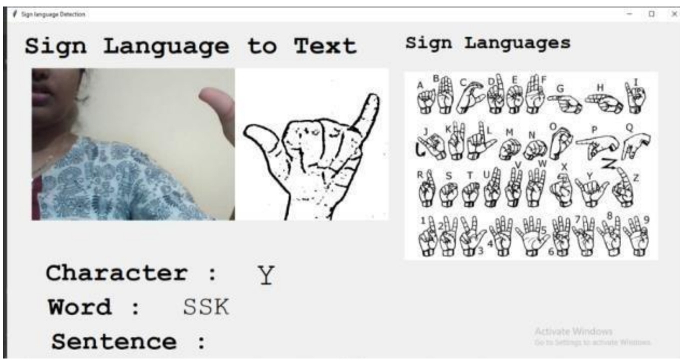

Figure 5. Sign Language of letter ‘Y’.

Figure 6. Blank for space in between the words.

In our code, we maintained the value as 50 and the distance threshold as 20, and displayed the letter whenever the count of a letter was detected above a certain number. Alternatively, to lessen the possibility that the projected letter would be inaccurate, we remove the current dictionary, which keeps track of how many times the current sign has been recognized. The absence of spaces is recognized if the current buffer is empty, and the number of blanks (plain background) detected is greater than a specific threshold. In the other scenario, it prints a space to indicatethe end of the word and appends the current to the phrase below. A blank is considered as space between two images as shown in Figure 6.

CONCLUSION

As a result, the gap in communication between the deaf and mute and others who could not comprehend sign language is closed. Its utilization will expand if it is made portable. We can use the equipment more fluidly and quicklythanks to the increased speed and precision. It can be used in situations where emergency communication is crucial, such as hospitals and police stations. It will assist in building a better society.

REFERENCES

1. Viswanathan, Shreyas, et al. "Sign Language To Text And Speech Conversion Using Cnn", volume-03, issue-05, May-2021, IRJMETS.
2. Deepika Bansal and Rita Chhikara, "An Improved DCT based Steganography Technique", International Journal of Computer Applications (0975– 8887), Volume 102– No.14, September 2014.
3. Adewale, Victoria A., and Adejoke O. Olamiti. "conversion of sign language to text and speech using machine learning techniques." (2018), Journal of research and review in science, Volume-5, December-2018, page-58-65.
4. NB, Mahesh Kumar. "Conversion of sign language into text." International Journal of Applied Engineering Research, Volume-13, issue-9, 2018, p-7154-7161, https://www.ripublication.com/ijaer18/ijaerv13n9_90.pdf.
5. Lahoti, Sakshi, et al. "Android based american sign language recognition system with skin segmentation and SVM." 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 10-12th July,2018, Bengaluru, IEEE Xplore, https://ieeexplore.ieee.org/document/8493838.
6. Vedak, Omkar, et al. "Sign language interpreter using image processing and machine learning." International Research Journal of Engineering and Technology (IRJET) (2019), Volume-6, issue-4, April-2019.
7. Tolentino, Lean Karlo S., et al. "Static sign language recognition using deep learning." Int. J. Mach. Learn. Comput Volume-9, issue-6, December (2019): 821-827.
8. Sruthi, C. J., and A. Lijiya. "Signet: A deep learning based Indian sign language recognition system." 2019 International conference on communication and signal processing (ICCSP). IEEE, 04-06 April,2019, Chennai, India, https://ieeexplore.ieee.org/abstract/document/8698006.

9. Kunjumon, Jinsu, and Rajesh Kannan Megalingam. "Hand gesture recognition system for translating indian sign language into text and speech." 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 27-29 November, Tirunelveli, India, 2019, https://ieeexplore.ieee.org/document/8987762.
10. Liao, Yanqiu, et al. "Dynamic sign language recognition based on video sequence with BLSTM-3D residual networks." IEEE Access, Volume-7, (2019), p-38044-38054, https://ieeexplore.ieee.org/document/8667292.
11. Tiku, Kohsheen, et al. "Real-time conversion of sign language to text and speech." 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 15-17th July, Coimbatore 2020, https://ieeexplore.ieee.org/document/9182877.
12. Wu, Ronghui, et al. "Full-fiber auxetic-interlaced yarn sensor for sign-language translation glove assisted by artificial neural network."Nano-micro letters 14.1 (2022), Springer, 1- 14.
13. Kirpan, Priyanka, et al. "Wearable Glove that Translate Sign Language into Text and Speech.", Volume-11, Issue-4, April-2022, Impact Factor-8.118, IJIRSET, P-3571-3575.

