VIVA Institute of Technology
10th National Conference on Role of Eng
ineers in Nation Building – 2022 (NCRENB-2022)

VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION
ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

A review on the perception and recognition systems for interpreting sign languages used by deaf and mute

Suyash Koltharkar[1], Amit Gupta[2], Hemil Patel[3], Sunita Naik[4]
[1](*Department of Computer Engineering, VIVA Institute of Technology, India*)
[2](*Department of Computer Engineering, VIVA Institute of Technology, India*)
[3](*Department of Computer Engineering, VIVA Institute of Technology, India*)
[4](*Asst. Professor, Department of Computer Engineering, VIVA Institute of Technology, India*)

Abstract: *Extensive research conducted across several fields revealed that hearing impairment and inability to verbally communicate lead to inequality of opportunity, as well as problems even in everyday life. Despite being a very useful medium of communication for deaf and mute people, sign language has no meaning for someone who does not understand it. The identification of these hand gestures is done by one of the two methods. Static images are one method of identifying while dynamic gestures are another. After skimming through the previous research, many limitations were exposed. The existing systems offered high accuracy rates upon feeding static images but the accuracy dropped significantly when dynamic inputs were fed. Some systems achieved good accuracy rates when fed with dynamic input but their scope was inadequate. After analyzing these techniques and identifying their limitations, we conclude with several promising directions for future research.*

Keywords - *Impairment, Deaf and Mute, Gestures, Static, Dynamic, Accuracy*

1. Introduction

Living in a privileged world of intellectuals and witnessing revolutions in the technological fields it is essential to not overlook the responsibilities to utilize technology to contribute to the progress and development of the society at large. The ability to communicate is key for any individual to lead a normal life. According to the definitions presented, anyone who cannot hear at all or can only hear loud noises is considered hearing impaired. A person who is deaf or whose speech isn't understood by a listener of normal comprehension and hearing is taken into account to possess a speech disability. An individual who is unable to talk because of speech disorders is considered mute. Most candidates who are hearing disabled are also speech disabled. According to Statista [11], in 2018, over 5% of the world's population – 360 million people – have hearing loss (328 million adults and 32 million children).

2. Background

In the course of extensive research conducted in various domains, researchers found that hearing impairment and inability to verbally express oneself causes disadvantages in terms of equal opportunity as well as causes communication issues in general. Sign Language, although being a medium of communication for deaf and mute people, still has no meaning when conveyed to a person who does not understand sign language.

Many nations have their interpretations of sign languages. The identification of the character's gesture is done by one of two methods. The first is a glove-based method in which the person wears a pair of gloves while capturing hand movements [5]. The second is a visual technique, which is divided into static and dynamic recognition. Static systems deal with the two-dimensional representation of gestures, while dynamics is a live, real-time capture of gestures. And despite an accuracy of over 77% [1], wearing gloves is uncomfortable and cannot be used in rainy weather. They are not easy to transport because they require computers to use. Hence to overcome the shortcomings of static hand gestures, dynamic hand gestures are used for better results.

3. Review of Literature

A survey was done on the existing literature and products to find out their shortcomings and research gaps in their systems. This survey consisted of more than 15 literature papers wherein the most relevant ones are shortlisted. The review of literature has been divided into three main categories to simplify the process, clarify and determine the advantages and shortcomings of the current technologies being used and being implemented, these categories are:

1 System with Static input:

Daphne Tan, et.al [1] carried out a sign language gesture recognition system using three various techniques. This research has explored CNN (Convolutional neural network) to detect shapes that are performed by using American sign language and then compare accuracies of various models. The only difference between these three models is that it uses various forms of the same database i.e some with no filtration or some with filtrations. Firstly the model was trained using a normal dataset without any preprocessing on it and was observed to give accuracy of 77% and alphabet accuracy of 74%. Secondly, it was performed using skin masking. As the name suggests it is obtained by masking the original image and just fetching the required part which in this case is the hand of the user. Here the result obtained was 92% and alphabet accuracy of 72%. Lastly, it was trained using Sobel filtration which means a gradient-based method that looks for strong changes in the first derivative of an image. It gave an accuracy of 91% and an alphabet accuracy of 71%.

Mehreen Hurroo, et.al [2] implemented sign language recognition using CNN and computer vision. Here A, B, C, D, H, K, N, O, T, and Y alphabets dataset were used for training and testing the model. Since output classes decreased there was an increase in the accuracy of the model. Before prediction, images undergo various preprocessing such as gray scaling, masking, segmentation, and feature extraction later. This model was able to give an accuracy of 98% which is considerably higher than the previous one with the disadvantage that it uses fewer classes for output.

Abdul Kawsar, et.al [6] were able to achieve an accuracy of 97% by using a CNN model by implementing a Faster Convergence and Reduction of Overfitting in Numerical Hand Sign Recognition using DCNN. The process starts by entering pre-processed data into the input layer. The system contains four pairs of convolution layers. Each convolution layer is followed by a max-pooling layer. These layers are equipped with an activation function called exponential linear unit. These set convolution and max-pool layers are followed by batch normalization layers. Each ASL Numerical class is equipped with 500 images, totaling 5000 images for 10 numerical classes. CNN with Dropout model gives an accuracy of 98.00% and CNN with Batchnorm and Dropout model gives an accuracy of 98.50%.

Sakshi Lahoti, et.al [8] successfully executed an Android-based American Sign Language Recognition System with Skin Segmentation and SVM. Here a dataset of 36 symbols containing alphabets A to Y, numbers from 0-9, and a spacebar was created. Z was not included in the database as it required video capturing and frame partitioning. Each symbol is trained using 500 images. The training images are resized to 200X200 pixels. The black background is used to make skin segmentation and edge detection easy. The sign language recognition process includes three steps; hand gesture capture and skin segmentation, feature extraction, and classification using SVM. Skin segmentation is done by YCbCr color systems. Feature extraction is done using HOG (histogram of oriented gradients). Finally, the classification is done using Support vector machines (SVMs). The system achieved an accuracy of 89.57%, though the accuracy may vary with complex backgrounds.

Shadman Shahriar, et.al [9] provided a study on the American Sign Language (ASL) finger speller based on Skin segmentation and machine learning algorithms. They have presented an automatic human skin segmentation algorithm based on color information. A YCbCr color space is used because it is typically used when encoding video and provides effective use of chrominance information for transforming human skin colors. In the CbCr plane, skin color distribution is modeled as a bivariate normal distribution. The algorithm's performance is demonstrated by simulations performed on images depicting people of various ethnicities. The performance of the algorithm is illustrated by simulations performed on images depicting people of different ethnicities. Then the convolutional neural network (CNN) is used to extract features from the images and the deep learning method is used to train a classifier to recognize sign language.

2 System with Dynamic input:

Anup Nandy, et.al [3] administered the classification of Indian sign language in real-time. Here first data was preprocessed using gray scaling followed by masking and then applying a threshold. This makes the data more precise to undergo a model for training and prediction. Then the research focuses on two main classification techniques namely using Euclidean distance and using K-nearest neighbor.

Panakala Kumar, et.al [4] devised a Video-Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic. Here data is filtered and preprocessed along with the Gaussian filter which smoothens the frames. The model uses fuzzy logic and Fourier descriptors for classification. With the inclusion of 80 signs, the model gave a recognition rate of 94%.

Jayan Mistry, et.al [7] materialized An Approach to Sign Language Translation using the Intel RealSense Camera. Here a dataset consists of 26 gestures (alphabets) with 100 samples extracted from 10 participants. Data is gathered using the Intel RealSense F200 camera and the RealSense API. 75% of the instances are used for training while the remaining 25% is used for testing purposes. Preprocessing is done using techniques available in the sci-kit-learn library. StandardScaler, MaxAbsScaler, and Normalization techniques are used for preprocessing. A multilayer perceptron (MLP) with three hidden layers gave the best results so it was used for all further experiments. The combination of SVM and MaxAbsScaler reached a performance of 95.00% while the MLP came close with 92.10%.

3 Physical System:

Carlos Fiel, et.al [5] resolved a Design of Translator Glove for Deaf-Mute Alphabet. The design of the glove consisted of hardware components like resistive sensors and an accelerometer placed on the glove which uses a technique to detect the exact position of fingers and hand movements. Data acquisition is done from the sensors via C programming language to identify the alphabet letters formed with fingers. Bend sensor, Pressure sensor, Accelerometer, and Signal conditioning circuit are the components required for reading the data. Components included in the processing stage are Microcontroller PIC16F887 and program. The translation phase includes a display LCD, EMIC 2 Text to Speech Module #30016, and a horn.

Ching-Hua Chuan, et.al [10], utilized a compact and affordable 3d motion sensor to demonstrate an American sign language recognition system. A palm-sized leap motion sensor offers more portability and greater economic benefits than the cyber globe or Microsoft Kinect currently used in studios. The 26 letters of the English alphabet in American sign language were classified using k nearest neighbors and support vector machines based on the sensory data. In the test results, the k nearest neighbor achieved 72 78% and the support vector machine e achieved 79 83% average sort rates, respectively. They also provided detailed discussions on the parameter setting in machine learning methods and the accuracy of specific alphabet letters in this paper.

4. Analysis

Table 1 represents the detailed analysis of the previous research using a tabular view of the Techniques, Advantages, and Limitations mentioned in each paper. This simplifies the review process and provides a clear description of different types of technologies and methods previously used to recognize sign languages.

TABLE 1
ANALYSIS TABLE

Sr No.	Title	Summary	Advantages	Technology Used
1	Implementing Gesture Recognition in a Sign Language Learning Application. [1]	Implementation of gesture recognition using CNN model to detect shapes and represent them in sign language.	Out of all the three models, the one with skin masking gives higher accuracy	Convolution Neural Network, TensorFlow, and Keras
2	Sign Language Recognition System using Convolutional Neural Network and Computer Vision. [2]	Captured images through camera and preprocessed the data with HSV color algorithm. The processed data is then fed to the CNN model for classification.	Low computing power and gives a remarkable accuracy of above 90%	Web Camera, Convolutional Neural Network
3	Classification of Indian Sign Language in real-time. [3]	An algorithm has been followed to calculate edge orientations in the sequence of ISL gestures which would be recognized using Euclidean distances and K-nearest neighbor methods.	More accurate classification results.	K-nearest neighbor and Euclidean distance, Gaussian filter

4	A Video-Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic. [4]	A wavelet-based video segmentation technique is proposed which detects shapes of various hand signs. Shape features are extracted using Fourier descriptors. Gesture recognition is done using the Sugeno-type fuzzy inference system.	The work was accomplished by training a fuzzy inference system by using features obtained using DWT and Elliptical Fourier descriptors by 10 different signer videos for 80 signs with a recognition rate of 96%.	Takagi-Sugeno-Kang (TSK), MATLAB
5	Design of Translator Glove for Deaf-Mute Alphabet. [5]	Implementation of a glove that helps deaf/mute communicate by detecting the movement of fingers and translating the same into words and sending audio signals.	The glove uses a portable voice synthesizer and a microcontroller instead of a computer or a smartphone to improve portability.	Bend sensor, Pressure sensor, Accelerometer, Microcontroller PIC16F887
6	Faster Convergence and Reduction of Overfitting in Numerical Hand Sign Recognition using DCNN. [6]	A layerwise optimized neural network architecture is proposed where batch normalization contributes to faster convergence of training. Batch normalization forces each training batch toward zero mean and unit variance, leading to improved flow of gradients through the model and convergence in a shorter time.	The proposed method achieves 98.50% accuracy over the constructed ASL-Numerical data set. Increased accuracy in a shorter time is achieved by diminishing overfitting.	Batch normalization, DNN
7	An Approach to Sign Language Translation using the Intel RealSense Camera. [7]	An Intel RealSense camera is used for translating static manual American Sign Language gestures into text. The highest accuracy of 95% is achieved by a support vector machine with a scaling method, as well as the principal component analysis used for preprocessing.	Classification of American Sign Language with the IntelRealSense camera is feasible with high accuracy and speed. A support vector machine used together with pre-processing yielded the best results.	IntelRealSense camera, MaxAbsScaler, PCA
8	Android-based American Sign Language Recognition System with Skin Segmentation and SVM. [8]	An Android application that captures images and converts the ASL to text is implemented. The system achieves an accuracy of 89.54% when skin segmentation is done by using the YCbCr system and classification of signs is done by the SVM model.	HOG (histogram of oriented gradients) is used instead of SIFT and other descriptors as it is unaffected by photometric and geometric transformations.	YCbCr color systems, HOG, SVM
9	Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning. [9]	A Real-time ASL fingerspelling translator based on skin segmentation and machine learning algorithms. ImageNet with 25 layers is used for datasets. CNN is used for feature extraction from images and a deep learning classifier is used to recognize sign language.	As the system detects skin color, YCbCr color space is employed because it is typically used in video coding and provides effective use of chrominance information for modeling the human skin color	YCbCr color space, CNN, AlexNet

10	American Sign Language Recognition Using Leap Motion Sensor. [10]	An American Sign Language recognition system using a compact and 3D motion sensor is presented. K-nearest neighbor and support vector machine to classify the 26 letters of the English alphabet in American Sign Language is applied.	The palm-sized Leap Motion sensor provides a much more portable and economical solution. The experiment result shows the highest average classification rate of 72.78% and 79.83%	Palm-sized Leap Motion sensor, K-nearest neighbor, Support vector machine.

5. CONCLUSION

Considering static images and dynamic videos used as a dataset for predicting sign languages lack accuracy in real-time. Talking about the static images, it doesn't support word recognition and only limits itself to alphabet recognition. Models based on the same were unable to predict words/alphabets which consisted of motion/gesture. On the other hand, video feed [4] doesn't support the formation of sentences which eventually breaks the flow of communication. Also, existing research focuses on specific language and sign language and there is no system that focuses on the facial expressions of a person because mainly they feed small frames of an image (only fist) to the model. Moreover, none of the existing research is open source. For a person who is deaf and mute, communication has always been an obstacle as it becomes difficult for a person with normal comprehension to interpret the hand signs. Also, sign language interpretations differ across the world. This increases the communication gap to a further extent. Thus there is a need for a system that will work on a dynamic video feed and will be able to form sentences with higher accuracy in a real-time environment.

6. REFERENCES

Proceedings Papers:

[1] Daphne Tan, Kevin Meehan, "Implementing Gesture Recognition in a Sign Language Learning Application" *IEEE, August 2020, 31st Irish Signals and Systems Conference (ISSC), 11-12 June 2020, Letterkenny, Ireland, DOI: 10.1109/ISSC49989.2020.9180197*

[2] Mehreen Murroo, Mohammad Elham Walizad, "Sign Language Recognition System using Convolutional Neural Network and Computer Vision", *IJERT, Vol. 9 Issue 12, 11 Dec. 2020, Paper ID : IJERTV9IS120029*

[3] Anup Nandy, Pavan Chakraborty, Jay Shankar Prasad, G. C. Nandi and Soumik Mondal, "Classification of Indian Sign Language in real time", *February 2010, Researchgate*

[5] Carlos Pesqueira Fiel, Cesar Cota Castro, Victor Velarde Arvizu, Nun Pitalúa-Díaz, Diego Seuret Jiménez, Ricardo Rodríguez Carvajal, "Design of Translator Glove for Deaf-Mute Alphabet", *3rd International Conference on Electric and Electronics (EEIC 2013), December 2013, DOI: 10.2991/eeic-13.2013.114*

[6] Abdul Kawsar Tushar, Akm Ashiquzzaman, Md. Rashedul Islam, "Faster Convergence and Reduction of Overfitting in Numerical Hand Sign Recognition using DCNN", *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 21-23 Dec. 2017, Dhaka, Bangladesh, DOI: 10.1109/R10-HTC.2017.8289040*

[7] Jayan Mistry, Benjamin Inden, "An Approach to Sign Language Translation using the Intel RealSense Camera", *2018 10th Computer Science and Electronic Engineering (CEEC), 19-21 Sept. 2018, Colchester, UK, DOI: 10.1109/CEEC.2018.8674227*

[8] Sakshi Lahoti, Shaily Kayal, Sakshi Kumbhare, Ishani Suradkar, Vikul Pawar, "Android based American Sign Language Recognition System with Skin Segmentation and SVM", *2018 9th International Conference*

on Computing, Communication and Networking Technologies (ICCCNT), 10-12 July 2018, Bengaluru, India, DOI: 10.1109/ICCCNT.2018.8493838

[9] Shadman Shahriar, Ashraf Siddiquee, Tanveerul Islam, Abesh Ghosh, Asir Intisar Khan, Celia Shahnaz, “Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning”, *TENCON 2018 - 2018 IEEE Region 10 Conference, 28-31 Oct. 2018, Jeju, Korea (South), DOI: 10.1109/TENCON.2018.8650524*

[10] Ching-Hua Chuan, Eric Regina, Caroline Guardino , “American Sign Language Recognition Using Leap Motion Sensor”, *2014 13th International Conference on Machine Learning and Applications, 3-6 Dec. 2014, Detroit, MI, USA, DOI: 10.1109/ICMLA.2014.110*

[11] Sevgi Z. Gurbuz, Ali Cafer Gurbuz, Evie A. Malaia, Darrin J. Griffin, Chris S. Crawford, Mohammad Mahbubur Rahman and Emre Kurtoglu, *“American Sign Language Recognition Using RF Sensing”, IEEE Sensors Journal, Volume: 21, Issue: 3, Feb.1, 1 2021, pp. 3763 - 3775.*

Journal Papers:

[4] P.V.V. Kishore, Panakala Rajesh Kumar, “A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic”, *International Journal of Engineering and Technology, January 2012, DOI: 10.7763/IJET.2012.V4.427*

[12] Sanjana R, Sai Gagana V, Vedhavathi K R, Kiran K N, “Video Summarization using NLP”, *International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 08, Aug 2021, pp. 3672-3675.*

[13] Huy B.D Nguyen and Hung Ngoc Do, “Deep Learning for American Sign Language Fingerspelling Recognition System”, *26th International Conference on Telecommunications (ICT), 15 August 2019,pp. 314-318*

[14] Huy B.D Nguyen and Hung Ngoc Do, “Deep Learning for American Sign Language Fingerspelling Recognition System”, *26th International Conference on Telecommunications (ICT), 15 August 2019,pp. 314-318*

[15] Rajesh George Rajan and M. Judith Leo, “American Sign Language Alphabets Recognition using Hand Crafted and Deep Learning Features”, *2020 International Conference on Inventive Computation Technologies (ICICT), 09 June 2020, pp. 430-434*