IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Survey on Sign Language Recognition Systems

[1]Shruty M. Tomar, [2]Dr. Narendra M. Patel, [3]Dr. Darshak G. Thakore

[1]M.Tech. Student, [2]Professor, [3]Professor and Head

[1]Computer Department,

[1]BVM Engineering College, V.V. Nagar, India

Abstract: Sign language is used by speech impaired people for communication but normal people do not know sign language which causes a communication gap between them. This gap can be filled using today's advanced technologies. Technologies such as image processing and machine learning can be used to build a system which converts sign language to text or speech. These systems can be of great help to dumb people as they can easily communicate with anyone using such system. This paper provides a brief survey of various research works carried out so far in this field.

Index Terms - Sign Language Recognition, Convolutional Neural Network.

I. INTRODUCTION

Speech impaired people face problems in communicating with people around them. Their family members and close friends learn sign language to communicate with them but other than that, normally, people do not know sign language and don't even try to learn, as they do not need it in their day-to-day life. The society does not understand how difficult it is for them to communicate with someone who does not know sign language.

Sign language is a language in which communication between people are made by visually transmitting the sign patterns to express meaning [2]. Sign language also includes facial expression and body language, these are known as non-manual signs and they play a major role in understanding correct meaning of signs [4]. Manual signs are the hand patterns and it contains the main information while non-manual signs are important for clarification [4]. Speech impaired people learn sign language so they can communicate with people around them and fulfil their day-to-day needs. Sign languages also have variations.

Different countries have their own sign language, for example American Sign Language (ASL), Argentinean Sign Language (LSA), British Sign Language (BSL), Indian Sign Language (ISL) etc [3]. Speech impaired people prefer sign language which is used in their region [3]. To bridge the communication gap between speech impaired people and normal people, a sign language recognition system can be build that converts sign language to text or speech using existing technologies.

Sign language recognition can be done by two methods. First one is sensor based technique, in this technique a sensor based wearable gadgets are used such as gloves [4]. This technique has two disadvantages, first one is that the gadgets are expensive and second these gadgets need to be carried everywhere otherwise the system would not work. Second approach involves computer vision based methods [4]. In these methods, camera is used to capture images for recognition. The second approach is more convenient to use as it does not need the user to wear any gadgets.

Many research work has been done on sign language recognition but most of the work is done on American sign language. Other languages are not explored as much as ASL. ASL is a single handed sign language, which makes it easier to work with, whereas in ISL some signs are performed with one hand and some signs are performed using both the hands which makes it complex to work with.

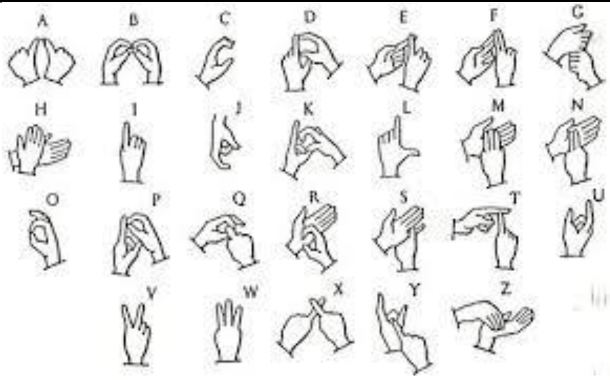

Figure 1: Indian Sign Languages Alphabets [12]

II. Literature Survey

Tanuj Bohra et al. [1] proposed a real-time two-way sign language communication system built using image processing, deep learning and computer vision. Techniques such as hand detection, skin color segmentation, median blur and contour detection are performed on images in the dataset for better results. CNN model trained with a large dataset for 40 classes and was able to predict 17600 test images in 14 seconds with an accuracy of 99%.

Joyeeta Singha and Karen Das [2] proposed a system for indian sign language recognition from a live video. The system comprises of three stages. Preprocessing stage includes skin filtering and histogram matching. Eigen values and eigen vectors are being considered for feature extraction stage and Eigen value weighted euclidean distance for classification. Dataset consisted 480 images of 24 signs of ISL signed by 20 people. System was tested on 20 videos and achieved an accuracy of 96.25%.

Muthu Mariappan H. and Dr. Gomathi V. [3] designed a real time sign language recognition system as a portable unit using contour detection and fuzzy c-means algorithm. Contours are used for detecting face, left and right hand. While fuzzy c-means algorithm is used to partition the input data into specified number of clusters. System was implemented on a dataset that contained videos recorded from 10 signers for several words and sentences. It was able to achieve accuracy of 75%.

Salma Hayani et al. [4] proposed an arab sign language recognition system based on CNN, inspired from LeNet-5 [13]. Dataset contained 7869 images of arab signs of numbers and letters. Various experiments were conducted by varying the number of training sets from 50% to 80%. 90% accuracy was obtained with 80% training dataset. The author has also compared the results obtained with machine learning algorithms like KNN (k-nearest neighbour) and SVM (support vector machine) to show performance of the system. This model was purely image based and it can be extended to video based recognition.

Kshitij Bantupalli and Ying Xie [5] worked on american sign language recognition system which works on video sequences based on CNN, LSTM and RNN. A CNN model named Inception was used to extract spatial features from frames, LSTM for longer time dependencies and RNN to extract temporal features. Various experiments were conducted with varying sample sizes and dataset consists of 100 different signs performed by 5 signers and maximum accuracy of 93% was obtained. Sequence is then fed to a LSTM for longer time dependencies. Outputs of softmax layer and maxpooling layer are fed to RNN architecture to extract temporal features from softmax layer.

Mahesh Kumar [6] proposed a system which can recognize 26 hand gestures of indian sign language based on Linear Discriminant Analysis (LDA) [14]. Preprocessing steps such as skin segmentation and morphological operations are applied on the dataset. Skin segmentation was carried out using otsu algorithm. Linear discriminant analysis is used for feature extraction. Each gesture is represented as a column vector in training phase which is then normalized with respect to average gesture. The algorithm finds the eigenvectors of the covariance matrix of normalized gestures. In recognition phase, subject vector is normalized with respect to average gesture and then projected onto gesture space using eigenvector matrix. Euclidean distance is computed between this projection and all known projections. The minimum value of these comparisons is selected.

Suharjito et al. [7] tried to implement a sign language recognition system with I3D inception [16] model through transfer learning method. Public dataset LSA64 [15] was used for 10 vocabularies with 500 videos. For training the dataset is distributed to 6:2:2 ratio, 300 videos for training, 100 for validation and 100 for testing set. The model has good training accuracy but very low validation accuracy.

Oscar Kellar et al. [8] introduced a hybrid CNN-HMM for sign language recognition. They conducted experiments on three datasets namely RWTH-PHOENIX-Weather 2012 [19], RWTH-PHOENIX-Weather Multisigner 2014 [20] and SIGNUM single signer [21]. Training and validation set have a ratio of 10 to 1. After the CNN training is finished a softmax layer is added and results are used in HMM as observation probabilities.

Mengyi Xie and Xin Ma [9] proposed an end-to-end system using residual neural network to implement recognition of american sign language. Dataset contained 2524 images for 36 classes. Data enhancement is used to expand the dataset to 17640 images. These images are converted to CSV file format and after applying one-hot encoding are given as input to ResNet50 [17] network for training. Model gives an accuracy of 96.02% without data enhancement and accuracy improves with data enhancement to 99.4%.

G. Anantha Rao et al. [10] proposes an indian sign language gesture recognition using convolutional neural network. This system works on videos captured from a mobile's front camera. Dataset is created manually for 200 ISL signs. CNN training is performed with 3 different datasets. In the first batch, dataset of only one set is given as input. Second batch has 2 sets of training data and third batch respectively has 3 sets of training data. Average recognition rate of this CNN model is 92.88%.

Aditya Das et al. [11] trained a convolutional neural network using Inception v3 [18] model for american sign language. Data augmentation is applied on the images before training them to avoid overfitting. This model gives more than 90% accuracy on Sreehari sreejith dataset [22] for 24 class labels with 100 images per class.

III. Comparison of Various Approaches

Most important requirement for sign language recognition system is the dataset. Datasets for languages such as american and arabic sign language are easily available on the internet but for other languages like indian sign language, any particular dataset is not available online, so a manual dataset has been developed by all the authors who have worked on it. Data augmentation can be carried out on small datasets in order to create variation of images that eventually expands the dataset and makes the classification model more robust.

Sign language recognition system contains various steps. First step is pre-processing of the dataset. Basic steps in pre-processing are resizing the image and conversion of image to grayscale or HSV or YCbCr [1][2][3]. Most important step in pre-processing is skin segmentation part, this step separates the skin portion from the non-skin portion in the image [1][2][3][6]. Skin segmentation can be done using OpenCV library in python or using otsu algorithm [1][2][3][6]. Morphological operations are applied on the resultant images for noise reduction [3][6]. Median blur can also be applied for smoothing of images [1].

Second step is feature extraction. Objective of feature extraction is to improve or maintain the accuracy of the classifier and simplify the complexity of the classifier [3]. Features can be extracted with the help of contours [1][3]. Features can also be extracted using other methods such as Eigen value and Eigen vectors and linear discriminant analysis [2][6]. LDA is used for dimensionality reduction and it finds a linear combination of features that classifies two or more classes of objects or events [6]. Extracted features are fed to neural network for classification.

In a normal neural network, weight updation can be infeasible due to large number of neurons and convolutional neural network helps in overcoming this drawback by reducing the dimension of input pixels [1][11][10][8][4]. Higher accuracy can be obtained using CNN as compared to traditional classifiers such as KNN and SVM [4]. Inception model is a variation of CNN in which the layers (convolutional, pooling and softmax) are stacked parallel to each other instead of on top of each other which helps in reducing processing and computational cost [5][7][11]. LSTM supports longer time dependencies that solves vanishing gradient problem of RNN and gives higher accuracies on long sequence of data [5]. Residual neural network solves the problem of gradient dispersion that occurs while increasing the depth of the network [9]. Euclidean distance between input and images stored in database can be computed and the minimum distance represents the recognized class [2][6]. Fuzzy c-means (FCM) classify the input into the appropriate class based on the degree of membership to the cluster centres [3]. Though FCM is more efficient and reliable than other clustering algorithms it requires more computation time than others.

IV. Limitations and Challenges

Models described in the given literature give poor results if the dataset includes faces of signers as the model ends up training incorrect features, same problem occurs with the color of the background. These models also face problems if they are trained on color images and the skin tone in testing images differ from training images. While working with videos, the models take a lot of time to predict sign and the dumb people are habituated with sign language so their speed cannot be matched with these existing systems.

Non-manual approach could be more effective in recognition as it considers facial expressions along with hand gestures but it can increase the complexity of implementation because variation in facial expression and body language can be much higher compared to variation in just hand gestures of different people. Though sensor-based techniques also give more accurate results in recognition, they suffer in aspects of portability and affordability.

V. Conclusion

In this paper, a survey on sign language recognition is presented and various techniques have been studied and analysed for the same. In recognition process, segmentation plays a crucial part in which skin region is separated from the background which usually affects the recognition accuracy. Besides segmentation, classification also depends on the feature extraction techniques which performs dimensionality reduction and reduces the computation cost. Study of various classification techniques concludes that deep neural network (CNN, Inception model, LSTM) performs better than traditional classifiers such as KNN and SVM.

REFERENCES

[1] T. Bohra, S. Sompura, K. Parekh and P. Raut, "Real-Time Two Way Communication System for Speech and Hearing Impaired Using Computer Vision and Deep Learning," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 734-739, doi: 10.1109/ICSSIT46314.2019.8987908.

[2] Singha, Joyeeta & Das, Karen. (2013), "Recognition of Indian Sign Language in Live Video," International Journal of Computer Applications. 70. 10.5120/12174-7306.

[3] H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862125.

[4] S. Hayani, M. Benaddy, O. El Meslouhi and M. Kardouchi, "Arab Sign language Recognition with Convolutional Neural Networks," 2019 International Conference of Computer Science and Renewable Energies (ICCSRE), Agadir, Morocco, 2019, pp. 1-4, doi: 10.1109/ICCSRE.2019.8807586.

[5] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.

[6] Kumar Mahesh, "Conversion of Sign Language into Text," International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 7154-7161.

[7] Suharjito, Suharjito & Gunawan, Herman & Thiracitta, Narada & Nugroho, Ariadi. (2018). Sign Language Recognition Using Modified Convolutional Neural Network Model. 1-5. 10.1109/INAPR.2018.8627014.

[8] Koller, Oscar & Zargaran, Sepehr & Ney, Hermann & Bowden, Richard. (2016). Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. 10.5244/C.30.136.

[9] M. Xie and X. Ma, "End-to-End Residual Neural Network with Data Augmentation for Sign Language Recognition," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 1629-1633, doi: 10.1109/IAEAC47372.2019.8998073.

[10] G. A. Rao, K. Syamala, P. V. V. Kishore and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), Vijayawada, 2018, pp. 194-197, doi: 10.1109/SPACES.2018.8316344.

[11] A. Das, S. Gawde, K. Suratwala and D. Kalbande, "Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-6, doi: 10.1109/ICSCET.2018.8537248.

[12] Elizabth Cody, "Deaftravel", http://www.deaftravel.co.uk/signs.php?id=27

[13]Richmond Alake, "Understanding and Implementing LeNet-5 CNN Architecture", June 2020, https://towardsdatascience.com/understanding-and-implementing-lenet-5-cnn-architecture-deep-learning-a2d531ebc342

[14]Priyankur Sarkar, "What is LDA: Linear Discriminant Analysis for Machine Learning", September 2019, https://www.knowledgehut.com/blog/data-science/linear-discriminant-analysis-for-machine-learning

[15] Ronchetti, Franco and Quiroga, Facundo and Estrebou, Cesar and Lanzarini, Laura and Rosete, Alejandro, "LSA64: A Dataset of Argentinian Sign Language", XX II Congreso Argentino de Ciencias de la Computación (CACIC),2016.

[16] Madeline Schiappa, "Understanding the Backbone of Video Classification: The I3D Architecture", June 2020, https://towardsdatascience.com/understanding-the-backbone-of-video-classification-the-i3d-architecture-d4011391692

[17] Priya Dwivedi, "Understanding and Coding a ResNet in Keras", January 2019, https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33

[18] Sik-Ho Tsang, "Review: Inception-v3-1st Runner Up (Image Classification) in ILSVRC 2015", September 2018, https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c

[19] J. Forster, C. Schmidt, T. Hoyoux, O. Koller, U. Zelle, J. Piater, and H. Ney. RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus. In Language Resources and Evaluation (LREC), pages 3785-3789, Istanbul, Turkey, May 2012.

[20] O. Koller, J. Forster, and H. Ney. Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding, volume 141, pages 108-125, December 2015.

[21] "Signum Database for Signer-Independent Continuous Sign Language Recognition", https://www.phonetik.uni-muenchen.de/forschung/Bas/SIGNUM/

[22] Sreehari Sreejith, "Weekend project: sign language and static-gesture recognition using scikit-learn", December 2016, Weekend project: sign language and static-gesture recognition using scikit-learn (freecodecamp.org)