Review on Sign Language Detection Using Machine Learning

Ms. Anamika Srivastava[1], Mr. Vikrant Malik[2]

1Dept. of Computer Science and Engineering

[2]Dept. of Information Technology Noida Institute of Engineering and Technology, Greater Noida, Uttar Pradesh

Email Id-[1] anamika@niet.co.in,[2] vikrantmalik.10@gmail.com

Received: 6 February 2020 Revised and Accepted: 21 May 2020

ABSTRACT: Sign language is the most popular and effective way for communication among hard hearing and normal people. Normal people find little difficulty in understanding and interpreting the meaning of sign language expressed by the hearing impaired, it is inevitable to have an interpreter for the translation of sign language. There are very fewer technologies that help to connect this social group to the world. Understanding sign language is the primary enablers in helping hard of hearing people with the rest of society. American Sign Language detection is a process in which computer analyses the American Sign Language gestures and then convert them into human-readable text. Those people who face difficulty in speaking and in hearing can easily communicate with the use of this Sign language detection software. Nowadays many types of research are going on to make this process easy and accurate. In this paper, an effort has been made to highlight the work and comparative study of that work done by researchers in American Sign Language.

KEYWORDS: American Sign Language, Deep learning, PCANet, English Sign Language, Microsoft Kinect,

I. INTRODUCTION

Language is a medium to communicate with a person or with a group. The spoken language is communication media for those who can speak and listen. The world has many more oral languages in different countries. Sign Language (SL) is a method of communication for the people who face problems in speaking and hearing. Those persons can communicate with each other or with the group by different signs and gestures. Many countries in the world have their style of Sign Language. For example American Sign Language (ASL), Indian Sign Language (ISL), British Sign Language (BSL), French Sign Language (FSL), Chinese Sign Language (CSL), Malaysian Sign Language (MSL) and many more.

ASL has been supposed a complete but difficult language, which uses signs with the moving hands and facial expressions. ASL has been mostly practiced and was considered as a key language in many North Americans with a hearing problem and many countries those who do not have their own Sign Language. One of the major problem faced by a person who is unable to speak is they cannot express their emotion as freely as they want. Utilize that voice recognition and voice search systems in smartphone(s)[1]. Audio results cannot be retrieved. They are not able to utilize (Artificial Intelligence/personal Butler) like Google assistance, or Apple's SIRI etc. because these work on voice controlling. There is a need for such platforms for such kind of people. American Sign Language (ASL) is a complete, complex language that uses signs made by moving the hands, facial expressions and postures of the body. It is the go-to language of many North Americans who are not able to talk and is one of the various communication alternatives used by people who are deaf or hard-of-hearing [2].

II. METHODOLOGY

Sign language is a method that individuals with loss in hearing and voice can interact. Individuals use expressions in sign language to communicate their feelings and desires through non-verbal contact. Nevertheless, it is incredibly difficult for non-signers to comprehend, which is why qualified sign language interpreters are needed for medical and legal activities, training and instructional sessions. The need for translating facilities has increased over the past five years. Many methods have been added, including video remote analysis and high-speed

Broadband connectivity. They also have a simple to use sign language communication tool that can be utilized, but which has major limitations such as internet connectivity and a compatible computer. One recommended modification is to check the experiment with further measures to assess the exactness of the measurements of greater sample scales and evaluate two separate CNN outputs. Another innovation is to use modern technology to measure results and see if the model will do better.

An analysis of the issue reveals that a variety of techniques have been used in video to tackle gesture detection using different methods. One communication used secret markov models (HMM), together with bayesian network classifiers and gaussian trees, to distinguish facial expressions from the video sequences. A paper on perception of human pose in a video series was also published in French by Francois using 2 D and 3 D approaches. The research states that silhouettes from a static camera are remembered by PCA and 3D are used as a picture location for recognition. This strategy has the downside of indirect movements which can contribute to training uncertainty and therefore lower predictability.

Let's address the study of video clips using neural networks, where visual knowledge is collected in the form of object vectors. Neural networks are associated with concerns such as hand-tracking, context and atmosphere segmentation, illumination, change, occlusion, orientation and location. The paper splits the dataset into parts, separates features and splits them into Euclidean and K-nearest.

White research describes how the Indian sign language should be understood on a continuous basis. The paper includes frame extraction from video files, files preprocessing, main frames extracted from the videos, and other functionality extracted, understood and eventually configured. The recording is transformed into RGB frames by preprocessing.

Every frame has the same scale. The segmentation of skin colors is used to separate skin regions through gradient AHS.

Such pictures have been converted into binary pictures. Through measuring a differential between the plates, food keyframes have been derived. And characteristics were derived using a histogram from the keyframes. Euclidean width, Manhattan duration, chess board duration, and Mahalanobis length were listed. CNN's or ConvNets are a category of neural networks which are respectable in the field of image recognition and classification. [9]CNN's use multilayer perceptron's which require minimal preprocessing to "train" the architecture to perform the task of recognition/classification very effective. [10]CNN's were modelled to perform like biological processes in terms of connectivity patterns between neurons in the visual cortex of animals. CNN's tend to perform better than other image and video recognition algorithms in fields of image classification, medical image analysis and natural language processing.

Recognition with Convolutional Neural Networks:

In a recent study, ASL letter classification done using convolutional neural network, that has showed incredible success in handling a variety of tasks related to processing videos and images.

The dataset contains 65000 images. All the images were colored in this experiment and height width ratio vary significantly but average approximately 150x150 pixels[3]. The hands have tightly cropped with no negative space and placed on a uniform black background. They employed Caffe, a deep learning framework, to develop, test, and run the CNNs. Berkeley Vision and Learning Center's GoogLeNet pre-trained used on the 2012 ILSVRC dataset. They attained a validation accuracy of nearly 98% with five letters and 74% with ten[4].

American Sign Language alphabet using depth images:

In this study, they propose a new user independent recognition system for American Sign Language alphabet using depth images, the images are captured from the low-cost Microsoft Kinect depth sensor. This overcomes many problems due to their robustness against illumination and background variations. Image acquisition is done using Microsoft Kinect, feature extraction is performed using Principal Component Analysis Network (PCANet) and classifying the data using support vector machine[5]. This system is tested using a public benchmark dataset collected from five users and gave average accuracy of 88.7%[6].

LeNet was one of the very first of CNN to be a pioneer in the areas of Multi-layer Perceptron and CNN analysis which led the way for further studies. This groundbreaking research by Yann LeCun, which had long since been popular since 1988, was called LeNet5. LeNet has primarily been developed for the identification of character tasks including digits and zip codes. From then on, the MNIST dataset has been developed to check the accuracy of each new neural network design proposed.

Connectivity is an issue to consider when dealing with high-dimensional inputs such as images, because connecting all the neurons with previous volumes does not take spatial structure into account. CNN's take advantage of local connection between neurons of nearby layers, the extent of which is a hyperparameter called receptive field. The connects are always in local in space, but they extend to the depth of input volume. Free parameters are controlled in convolutional layers by using the concept of parameter sharing. It relies on the assumption that a patch feature is reusable and can be used in different layers of the neural network. The accuracy of misclassified signs did not correct with an increase in sample size, in fact originally correctly classified signs were later misclassified when increasing the number of signs which leads to a conclusion that there could possibly be too little difference between those signs for the model to differentiate and this paper need more features or more distinction to have better accuracy.

In this paper this paper introduced a way to recognize American Sign Language using machine learning. It is an approach to solve the problems faced by people with hearing and speech impairments. It's composed of 2 major components, analyzing the gestures from images and classifying images. Since this paper is dealing with a smaller dataset, using a larger dataset may provide better results. This paper investigated two approaches to classification: using the pool layer and using the SoftMax layer for final predictions. The SoftMax layer provided better results because of distinct features. The sheer number of features in a 2048 vector confused the network leading to poorer results.

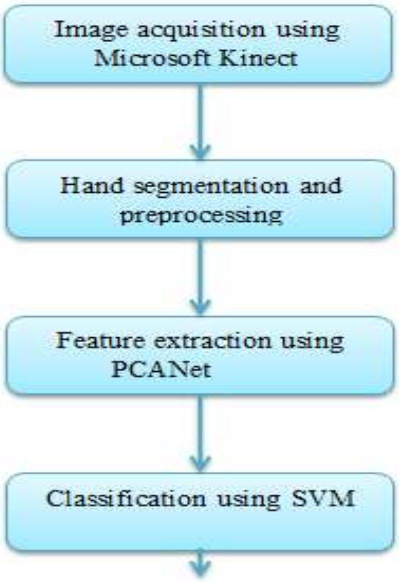
Image acquisition using Microsoft Kinect
Hand segmentation and preprocessing
Feature extraction using PCANet
Classification using SVM

Fig. 1: Proposed Method for American Fingerspelling Recognition

Deep Learning on Custom Processed Static Gesture Images:

In this research, the dataset has images of American Sign Language and there are 24 labels of static gestures from letters A to Y, excluding J, There are on an average 100 images per class. From each of the classes, 20% of the images were used specifically for testing. Data Augmentation is done by cropping, scaling, rotating and flipping images, and classification is performed by convolutional neural network model using Inception v3[7] . Testing of images on the trained model to determine the level of accuracy of the final model has been implemented using the Tensor Flow python library. They got an average validation accuracy of 90% with the greatest validation accuracy being 98%[8] as shown in fig.1 and 2.

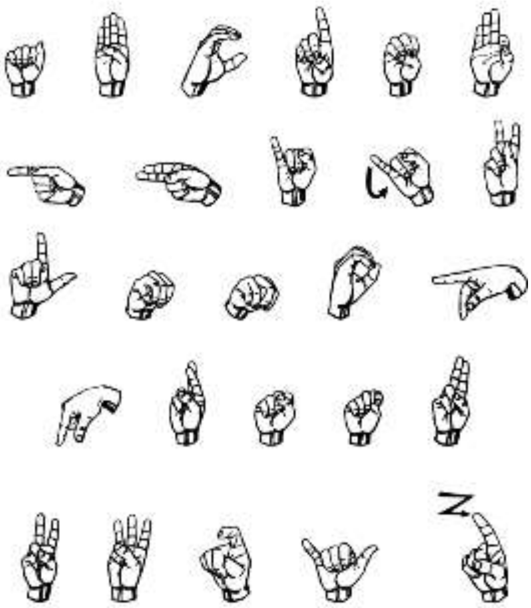

Fig. 2: Basic Gestures in American Sign Language

Sign Language Using Deep Neural Network:

In this paper, Images of hand gestures for English Sign Language (ESL) is obtained from Kaggle website. A total of 810 images for the ESL are used for the training of the DNN. The dimension of each image is 200x200 pixels and it is reduced to standard form of 128x128 pixels. Deep Neural Network (DNN) based machine translation is used for feature extraction. The proposed system uses a three layer deep Convolutional Neural Network (CNN) for hand gesture recognition system. The input layer has the dimension of (3,128,128) and the first data specifies the RGB channel and other data specify the input image dimension. The first ConvNet block has 16 filters of size (5, 5) followed by a Max-Pooling layer of size (2, 2) [9]. This is followed by other two ConvNet layers. They got a peak accuracy of 100% for training process and 82% for validation process. Additionally, a test accuracy of 70% was obtained for the sample test images[10].

III. CONCLUSION

This review paper presents various algorithms, and techniques like Principal Component Analysis Network (PCANet), convolutional neural network, Inception v3, Most of the SL and gesture recognition problems have been addressed based on Statistical Modeling such as PCA, support vector machine are used for gesture recognition. From the above consideration it is clear that the convolutional neural network model using Inception v3 has made outstanding progress in the field of gesture recognition. With the average validation accuracy of 90%, with the greatest validation accuracy being 98%.

IV. REFERENCES

[1] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, “American sign language recognition with the kinect,” in ICMI’11 - Proceedings of the 2011 ACM International Conference on Multimodal Interaction, 2011, doi: 10.1145/2070481.2070532.

[2] C. Savur and F. Sahin, “American Sign Language Recognition system by using surface EMG signal,” in 2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings, 2017, doi: 10.1109/SMC.2016.7844675.

[3] L. Pigou, S. Dieleman, P. J. Kindermans, and B. Schrauwen, “Sign language recognition using convolutional neural networks,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015, doi: 10.1007/978-3-319-16178-5_40.

[4] G. A. Rao, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry, “Deep convolutional neural networks for sign language recognition,” in 2018 Conference on Signal Processing And Communication Engineering Systems, SPACES 2018, 2018, doi: 10.1109/SPACES.2018.8316344.

[5] S. Aly, B. Osman, W. Aly, and M. Saber, “Arabic sign language fingerspelling recognition from depth and intensity images,” in 2016 12th International Computer Engineering Conference, ICENCO 2016: Boundless Smart Societies, 2017, doi: 10.1109/ICENCO.2016.7856452.

[6] L. Zheng and B. Liang, “Sign language recognition using depth images,” in 2016 14th International Conference on Control, Automation, Robotics and Vision, ICARCV 2016, 2017, doi: 10.1109/ICARCV.2016.7838572.

[7] S. Agrawal, R. Rangnekar, A. Das, S. Gawde, and S. Dhage, “Gauging Customer Interest Using Skeletal Tracking and Convolutional Neural Network,” in Proceedings of 2019 3rd IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2019, 2019, doi: 10.1109/ICECCT.2019.8869045.

[8] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, doi: 10.1109/CVPR.2017.195.

[9] P. T. Krishnan and P. Balasubramanian, “Detection of Alphabets for Machine Translation of Sign Language Using Deep Neural Net,” in 2019 International Conference on Data Science and Communication, IconDSC 2019, 2019, doi: 10.1109/IconDSC.2019.8816988.

[10] F. Monti, D. Boscaini, J. Masci, E. Rodolà, J. Svoboda, and M. M. Bronstein, “Geometric deep learning on graphs and manifolds using mixture model CNNs,” in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, doi: 10.1109/CVPR.2017.576.