A Software based ASL Translator and ASL to Voice Emulation System

Joel D[1], Karthika Rajeev[2], Lenin Ben[3], Jerome Pathrose[4], Ms.Sujitha M[5]
[1]Mangalam College of Engineering, India, joelggez@gmail.com
[2] Mangalam College of Engineering, India, karthikarajeevm@gmail.com
[3]Mangalam College of Engineering, India, leninben12345@gmail.com
[4]Mangalam College of Engineering, India, jeromepathrose97@gmail.com
[5]Faculty Mangalam College of Engineering, India, m.sujitha@mangalam.in

ABSTRACT

Sign languages use the visual modality to convey meaning. People who are hearing and speech impaired are left behind video consultations as there is no easier way for non vocal people to communicate with vocal people and express themselves to the hearing world. Since they can't sign themselves they are forced to use a basic text chat to hold their consults. To overcome this issue, we have developed a real time software based ASL Translator and voice emulation system where the image captured is converted into predicted text. We use Deep Neural Networks like Convolutional Neural Network for performing feature extraction and the SVM classifier for classification. The system returns top and most probable predictions. We obtained an accuracy rate of 93.75%.

Key words: ASL, Convolutional Neural Networks, Deep Neural Networks, HOG

1. INTRODUCTION

Sign Languages are languages that are mainly expressed through a process called signing, this can be further explained as the usage of visual-manual modality to convey an idea or meaning with specificity. The main limitation is the fact that sign languages are not universal hence leading to the factor that each type of sign languages have their own grammar and lexicon. In all the communities where deaf people exist there is a need for communication between the speech impaired and the healthy ones. This gap between

the group can only be closed with the introduction of a means of communication between the two. The communication between the signers and non-signers would be a relevant issue worth addressing. As of the issues faced by the speech impaired community due to the fact that people who don't have the necessity to deal with deaf people, have no use of learning any sign language, hence resulting in a community of people with a majority of non-signers. This brings up the issue of a speech impaired person

communicating with a non-signer, which prevents them from expressing basic ideas to daily necessities with each other.

An ASL translator which converts the American Sign Language (ASL) to speech through a voice emulation system which in turn would solve the communication issues between signers and non-signers. It mainly focuses on translating the ASL finger spelled alphabets in a real time system with OpenCV by reading frames from a camera and classifying them frame by frame. The Pipelining can be explained as a five step process Data Collection, Pre-processing, Feature extraction, Machine learning and closes with a Real time system. Basically, it loads a script onto a pre-trained model with a custom classifier which classifies the ASL alphabets frame-by-frame in real time. The script causes the associated hardware camera to capture a live freed with a Region of Interest which is further cropped and processed and the classifier returns the top predictions made by the classifier.

2. RELATED WORKS

Acquisition of data and classification are the main procedures in Sign Language Recognition. According to [1,13], the lately captured picture is compared with images for specific letters in the database with the usage of the Scale Invariant Feature Transform (SIFT) algorithm and at the end of the comparison the gesture that was shown is delivered and the translated text for the following gesture. In [2,11,12], the authors develop a sign language recognition system using Convolution Neural Networks(CNN) where individual frames are extracted from the video captured through the camera and the frames extracted are subjected to gesture recognition. In [3], the authors make use of gloves for data acquisition where the wearable gloves connect hand gestures to an audio output and interpret the expression achieved by integrating a flex sensor and 8051 microcontroller. According to [4], the image obtained is advanced through a RetinaNet based hand detector to extract the hand regions and then they are detected, extracted and passed through a light - weight Convolutional Neural Network(CNN) for recognizing the hand gestures.. In [5,14], the Deep Belief Networks (DBN), 3D Convolutional Neural Network (CNN)

and Hidden Markov Model (HMM)are used where DBN and CNN are used for extracting skeletal and depth images. In [8,15], during the feature extraction localized contour sequences and block based features are extracted which helps in a much better characterisation of static hand gesture. In [9], the authors use a speeded up robust features algorithm for recognizing the static gestures by making use of the kinetic sensor and they focus on the Mandarin - Tibetan language. According to [10], Leave One Out Training and Testing strategy and an HMM based ensemble of classifiers is used for capturing gestures which is evaluated as a favourable strategy for gesture recognition.

3. PROPOSED SYSTEM

The ASL translator developed will convert the American Sign Language (ASL) to speech through a voice emulation system.The system will access your web camera and will open a window with the live camera and then the video is captured. This captured video is divided into several individual frames and they are converted into grayscale images and we clear it using the Gaussian blur. The edges of the hand are detected using the canny edge detection algorithm. We draw bounding boxes to separate the required and unwanted portion of the image where only the required portion is only used for further processing. We use pre trained Deep Neural Network(DNN) models like Convolutional Neural Networks (CNN) which are included in the keras for extracting the features.
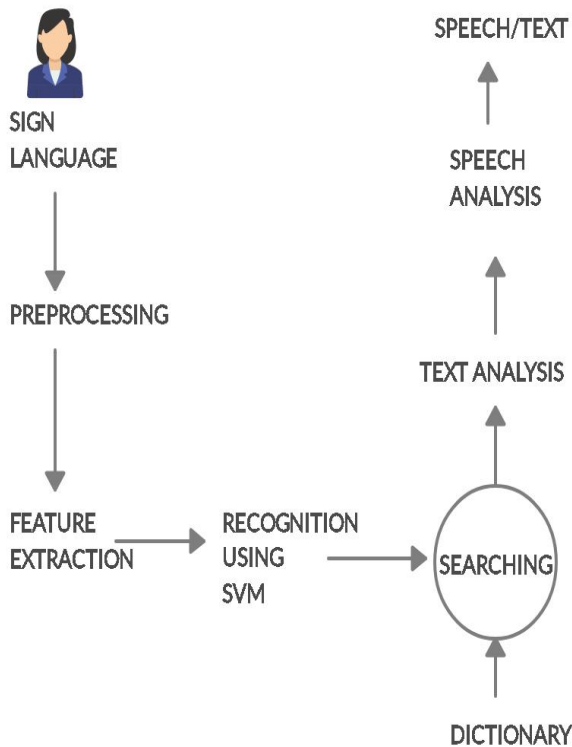
The Pipelining can be explained as a five step process Data Collection, Pre-processing, Feature extraction, Machine learning And closes with a Real time system. The script causes the associated hardware camera to capture a live freed with a Region of Interest which is further cropped and processed and the classifier returns the top three predictions made by the classifier. The top prediction is the letter with the highest percentage and the most probable predictions are also shown. Based on the classes the Support Vector Machine(SVM) is trained based on the feature descriptor Histogram of Oriented Gradients(HOG). Finally, after predicting the correct result we show the words or sentences based on the input from the user in real time.
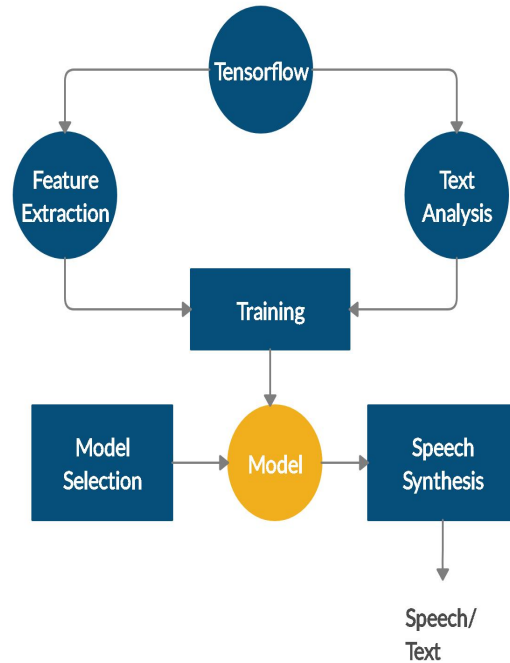

Figure 2: Training and Recognition

4. ALGORITHM

In this paper, we use Convolutional Neural Networks for implementing the system. We capture the images in real time using a webcam. After obtaining the images frame by frame we remove the background in order to avoid noises in the data. Then the image undergoes various filtering techniques and we create bounding boxes [7] around them. Finally using Convolutional Neural Networks we classify the data. The features are identified and they are matched with the data already present in the database. Based on the percentage of matching, the most suitable letter is predicted by the second and third accurate letters at the bottom.
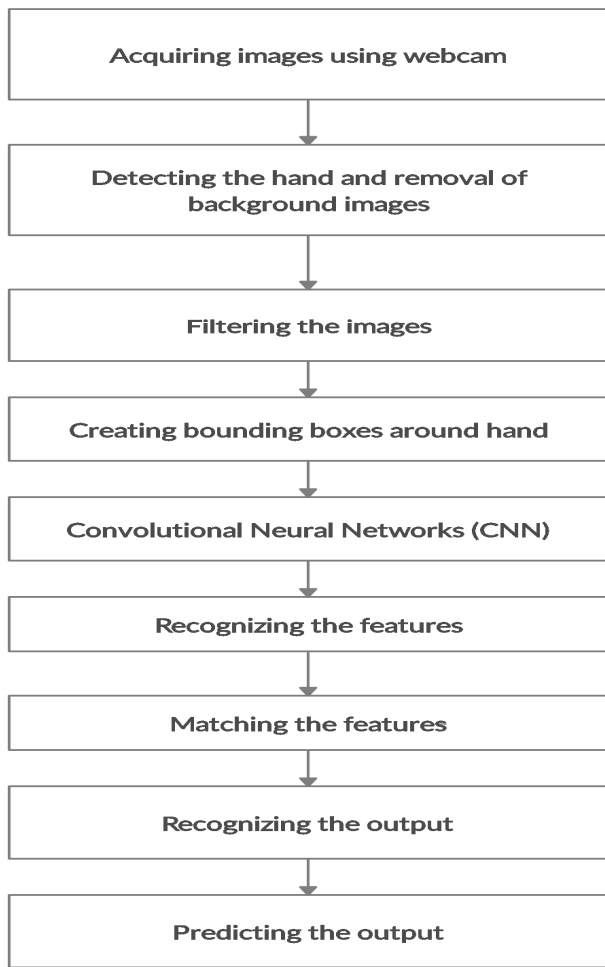

Figure 1: Sign Language Recognition

Figure 3: Steps involved in predicting the output

5. EXPERIMENT RESULTS

The most crucial procedure in developing a Sign Language Translator is the collection of sufficient data. The more number of data we have, the chances of us developing a good ASL Translation system with a better accuracy rate is quite high. The system was developed with python 3.7. For testing purpose libraries like OpenCV 3.4.0, Tensorflow GPU version, Keras 2.3.0, numpy 1.18.2, joblib 0.14.2, sklearn 0.22.2, pytt sx3 2.7 were used. For the development of the system we also require GPU above Nvidia GTX 1650 and a processor above Intel i7 9th generation or above.

The following figure shows the accuracy rate obtained for each letter. Letters K and R show a less accuracy rate when compared to other letters. We obtained a total accuracy rate of 93.75%. Fig 4 shows the output prediction of the system where the most suitable predictions are listed with their percentage of matching.
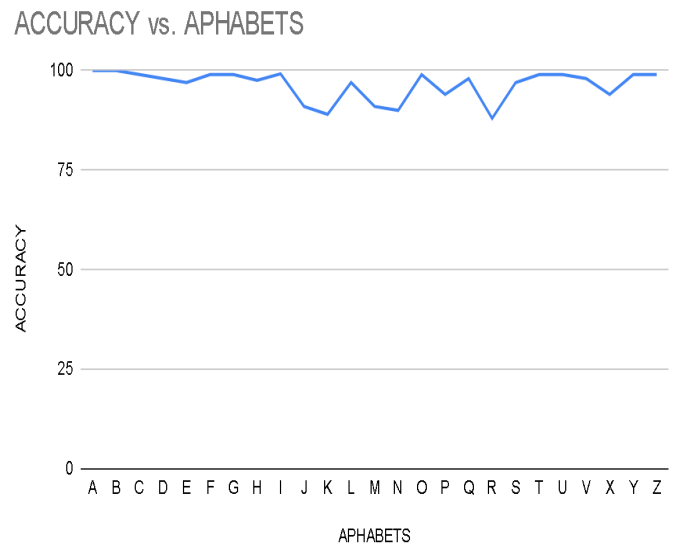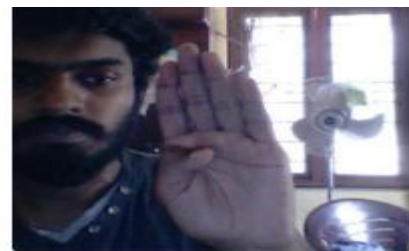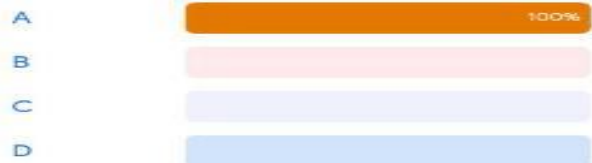

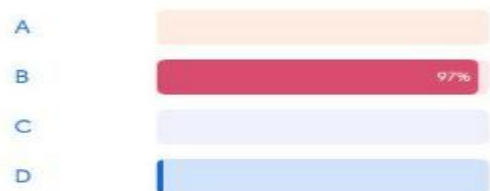Figure 4: Accuracy rate of ASL alphabets

Figure 5: Output Prediction

6. FUTURE SCOPE

In future, we can make this system further efficient by identifying emotional gestures via face recognition. Now we mainly focused on identifying the hand gestures. We can also make the system more efficient by including mobile phone camera along with web camera. Currently the system focuses only on American Sign Language. We can further expand the system by including other sign languages like Australian Sign Language(AUSLAN), Indian Sign Language(ISL). We can also implement word level and sentence level translation in future.

VII. CONCLUSION

In this paper, we developed an American Sign Language Translator and voice emulation system which helps to bridge the communication gap between normal and hearing or speech impaired people. In this system, we make use of a web camera to capture images in real time. Feature extraction is performed with the help of Deep Neural Networks like Convolutional Neural Networks.And the classifier classifies the output. We are provided with mainly three predictions including the top and most probable predictions. Here the translation is performed in real time which in turn will help the hearing and speech impaired people to communicate with each other. This system paves a way for the deaf and mute to communicate with the others effectively. The system provides an accuracy rate of 93.75%.

REFERENCES

[1] Er. Kanika Goyal1, Amitoj Singh, "Indian Sign Language Recognition System for Differently-abled People", *Journal on Today's Ideas – Tomorrow's Technologies*, Vol. 2, No. 2, pp. 145–151, 2014.

[2] L. Boppana, R. Ahamed, H. Rane, and R. K. Kodali, "Assistive Sign Language Converter for Deaf and Dumb", *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom), and IEEE Smart Data (SmartData)*, pp. 302-307, July 2019.

[3] D Jadhav and Dr. A.Tripathy, "Gesture Aided Speech for Deaf and Mute", 2018 3rd International Conference for Convergence in Technology (I2CT), pp. 1-6, April 2018.

[4] A. A. Q. MOHAMMED, J. Lv, and M. D. S. Islam, "A Deep Learning-Based End-to-End Composite System for Hand Detection and Gesture Recognition," *Sensors*, vol. 19, no. 23, p. 5282, Nov. 2019.

[5] Di Wu, Ling Shao, " Deep dynamic neural networks for gesture segmentation and recognition", *In Computer Vision-ECCV 2014 Workshops*, vol. 8925, pp. 552–571, March 2015.

[6] V. N. T. Truong, C. Yang and Q. Tran, "A translator for American sign language to text and speech," 2*016 IEEE 5th Global Conference on Consumer Electronics, Kyoto,* pp. 1-2, 2016.

[7] S. Shahriar *et al.,* "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*, Jeju, Korea (South), pp. 1168-1171, 2018.

[8] D. K. Ghosh and S. Ari, "Static Hand Gesture Recognition Using Mixture of Features and SVM Classifier," *2015 Fifth International Conference on Communication Systems and Network Technologies*, Gwalior, pp. 1094-1099, 2015.

[9] H. Yang, X. An, D. Pei and Y. Liu, "Towards realizing gesture-to-speech conversion with a HMM-based bilingual speech synthesis system," *2014 International Conference on Orange Technologies*, Xian, pp. 97-100, 2014.

[10] V. Godoy, A. S. Britto, A. Koerich, J. Facon and L. E. S. Oliveira, "An HMM-Based Gesture Recognition Method Trained on Few Samples," *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, Limassol, pp. 640-646, 2014.

[11] Vinodh P Vijayan, Deepti John, Merina Thomas, Neetha V Maliackal, Sara Sangeetha Varghese "Multi Agent Path Planning Approach to Dynamic Free Flight Environment", International Journal of Recent Trends in Engineering (IJRTE), ISSN 1797-9617 Volume 1, Number 1, May 2009, Page(s): 41-46.

[12] Juby Joseph, Vinodh P Vijayan" Misdirection Attack in WSN Due to Selfish Nodes; Detection and Suppression using Longer Path Protocol" International

Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 7, July 2014, pp. 825-829, ISSN: 2277 128X

[13] V P Vijayan, Biju Paul "Multi Objective Traffic Prediction Using Type-2 Fuzzy Logic and Ambient Intelligence" International Conference on Advances in Computer Engineering 2010, Published in IEEE Computer Society Proceedings, ISBN: 978-0-7695-4058-0, Print ISBN: 978-1-4244-7154-6

[14] Vijayan V P, Gopinathan E "Improving Network Coverage and Life-Time in a Cooperative Wireless mobile Sensor Network " Fourth International Conference on Advances in computing and communications (ICACC) Aug, 2014. Published in IEEE Computer Society Proceedings. Print ISBN: 978-1-4799-4364-7, INSPEC AccessionNumber:14630874,DOI:10.1109/ICACC.20 14.1 6 PP 42-45.

[15] Vinodh P Vijayan, Biju Paul " Traffic scheduling for Green city through energy efficient Wireless sensor Networks " International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.4, July – August 2019, ISSN 2278-3091, https://doi.org/10.30534/ijatcse/2019/81842019.