Together- Video Call For SLP (Sign Language People)

Sanjay J P[1], Saranya A[2], Sarulatha M[3], Lilly Florence M[4]
[1, 2, 3] Dept of Computer Science and Engineering
[4]Assistant Head of the Department, Dept of Computer Science and Engineering
[1,2,3,4] Adhiyamaan College of Engineering, Hosur, India

Abstract- *Sign languages are used by the people who can't speak and hear. So, to convey their message they show signs. As we already know these signs are only understandable when the other person also get to know the signs. In this case the sign language detection are used which predicts the sign. But this is possible when they are near and the distance communication is not possible. So to make interaction between sign language people and common language people even when they are far is by using video call. In this when the sign language people show the sign it will predict and show the message on the other side so that the person who can't speak sign language can be able to understand what they are saying through video call. The accuracy percent for the prediction of sign is 70%. Sign recognition system provides us an innovative, natural, user friendly way of interaction with the computer which is more familiar to human beings. Gesture Recognition has a wide area of application including human machine interaction, sign language, immersive game technology etc. By keeping in mind the similarities of human hand shape with four fingers and one thumb, this aims to present a real time system for hand gesture recognition on the basis of detection of some meaningful shape based features like orientation, center of mass (centroid), status of fingers, thumb in terms of raised or folded fingers of hand and their respective location in image. They predict the signs using CNN algorithm with already implemented supervised machine learning algorithms with respect to its accuracy, precision, sensitivity etc.*

Keywords- Sign language, Gesture, Recognition, RTCPeerconnection

I. INTRODUCTION

Interface with computers using gestures of the human body, typically hand movements. In gesture recognition technology, a camera reads the movements of the human body and communicates the data to a computer that uses the gestures as input to control devices or applications. One way gesture recognition is being used is to help the physically impaired to interact with computers, such as interpreting sign language. Unlike Haptic interfaces, gesture recognition does not require the user to wear any special equipment or attach any devices to the body. The gestures of the body are read by a camera instead of sensors attached to a device. Communication leads to better understanding, and it encompasses all the members of the community, including hearing impaired. Sign language bridges the gap of communication with other people. However, most hearing people do not understand sign language and learning it is not an easy process. As a result, there is still an undeniable barrier between the hearing impaired and hearing majority. Over the past few decades, many efforts have been made in creating a sign language recognition (SLR) system.

There are two main categories in SLR, namely isolated sign language recognition and continuous sign classification. American Sign Language (ASL) substantially facilitates communication in the hearing impaires community. The alternative of written communication is cumbersome, impersonal and even impractical when an emergency occurs. In order to diminish this obstacle and to enable dynamic communication, we present an ASL recognition system in video call application in which they can even communicate from long distance.

Sign recognition system provides us an innovative, natural, user friendly way of interaction with the computer which is more familiar to human beings.

The constraints imposed by the extra requirements reduce the scalability and feasibility of these solutions. This aims to make interaction between sign language people and common language people through video call website using webRTC and the RTCPeerConnection interface represents a WebRTC connection between the local computer and a remote peer where the models are trained using Deep learning concepts and predicts the sign.

II. LITERATURE REVIEW

American Sign Language (ASL) substantially facilitates communication in the deaf community. However, there are only ~250,000-500,000 speakers which significantly limits the number of people that they can easily communicate with [1]. The alternative of written communication is cumbersome, impersonal and even impractical when an emergency occurs. In order to diminish this obstacle and to enable dynamic communication, we present an ASL recognition system that uses Convolutional Neural Networks (CNN) in real time to translate a video of a user's ASL signs into text.

While Neural Networks have been applied to ASL letter recognition (Appendix A) in the past with accuracies that are consistently over 90% [2-11], many of them require a 3-D capture element with motion-tracking gloves or a Microsoft Kinect, and only one of them provides real-time classifications. The constraints imposed by the extra requirements reduce the scalability and feasibility of these solutions.

ASL recognition is not a new computer vision problem. Over the past two decades, researchers have used classifiers from a variety of categories that we can group roughly into linear classifiers, neural networks and Bayesian networks [211]. While linear classifiers are easy to work with because they are relatively simple models, they require sophisticated feature extraction and preprocessing methods to be successful [2, 3, 4]. Singha and Das obtained accuracy of 96% on 10 classes for images of gestures of one hand using Karhunen-Loeve Transforms [2]. These translate and rotate the axes to establish a new coordinate system based on the variance of the data. This transformation is applied after using a skin filter, hand cropping and edge detection on the images. They use a linear classifier to distinguish between hand gestures including thumbs up, index finger pointing left and right, and numbers (no ASL). Sharma et al. use piece-wise classifiers (Support Vector Machines and k-Nearest Neighbors) to characterize each color channel after background subtraction and noise removal [4]. Their innovation comes from using a contour trace, which is an efficient representation of hand contours. They attain an accuracy of 62.3% using an SVM on the segmented color channel model.

Bayesian networks like Hidden Markov Models have also achieved high accuracies [5, 6, 7]. These are particularly good at capturing temporal patterns, but they require clearly defined models that are defined prior to learning. Starner and Pentland used a Hidden Markov Model (HMM) and a 3-D glove that tracks hand movement [5]. Since the glove is able to obtain 3-D information from the hand regardless of spatial orientation, they were able to achieve an impressive accuracy of 99.2% on the test set. Their HMM uses timeseries data to track hand movements and classify based on where the hand has been in recent frames.

Suk et al. propose a method for recognizing hand gestures in a continuous video stream using a dynamic Bayesian network or DBN model [6]. They attempt to classify moving hand gestures, such as making a circle around the body or waving. They achieve an accuracy of over 99%, but it is worth noting that all gestures are markedly different from each other and that they are not American Sign Language. However, the motion-tracking feature would be relevant for classifying the dynamic letters of ASL: j and z.

Some neural networks have been used to tackle ASL translation [8, 9, 10, 11]. Arguably, the most significant advantage of neural networks is that they learn the most important classification features. However, they require considerably more time and data to train. To date, most have been relatively shallow. Mekala et al. classified video of ASL letters into text using advanced feature extraction and a 3-layer Neural Network [8]. They extracted features in two categories: hand position and movement. Prior to ASL classification, they identify the presence and location of 6 "points of interest" in the hand: each of the fingertips and the center of the palm. Mekala et al. also take Fourier Transforms of the images and identify what section of the frame the hand is located in. While they claim to be able to correctly classify 100% of images with this framework, there is no mention of whether this result was achieved in the training, validation or test set.

Admasu and Raimond classified Ethiopian Sign Language correctly in 98.5% of cases using a feed forward Neural Network [9]. They use a significant amount of image preprocessing, including image size normalization, image background subtraction, contrast adjustment, and image segmentation.

Admasu and Raimond extracted features with a Gabor Filter and Principal Component Analysis. The most relevant work to date is L. Pigou et al's application of CNN's to classify 20 Italian gestures from the ChaLearn 2014 Looking at People gesture spotting competition [11]. They use a Microsoft Kinect on fullbody images of people performing the gestures and achieve a crossvalidation accuracy of 91.7%. As in the case with the aforementioned 3-D glove, the Kinect allows capture of depth features, which aids significantly in classifying ASL sign.

III. EXISTING SYSTEM

The existing study aims to develop a system that will recognize static sign gestures and convert them into corresponding words. A vision-based approach using a web camera is introduced to obtain the data from the signer and can be used offline. The purpose of creating the system is that it will serve as the learning tool for those who want to know more about the basics of sign language such as alphabets, numbers, and common static signs.

DISADVANTAGES

- No Direct communication
- Low accuracy and predictions
- It could not be accurate
- Less stability provides
- Time consuming for capturing images

IV. PROPOSED SYSTEM

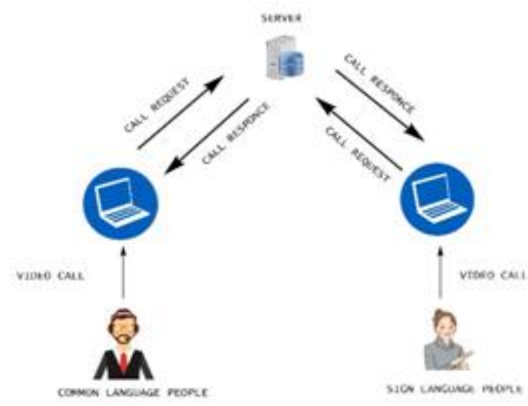
In proposed system the sign language people and common language people will be able to communicate through devices which makes the distance communication possible and easy. In this video call the actions of sign language people is predicted and shown for common language people that they can understand their message directly even on long distance. The camera will capture the images of the hands that will be fed in the system. Note that the signer will adjust to the size of the frame so that the system will be able to capture the orientation of the signer's hand. If the edges of the hand in the masking are detected clearly. Three trials were conducted in each letter/number/word gesture recognition.

ADVANTAGES

- Communication is possible through devices.
- Can understand the message easily
- More efficient and feasible

VI. CONCLUSION

The existing system we can only understand what they are saying when they know it or by using the sign detector. But the long-distance communicationis not possible. So, it is has become a disadvantage for the sign language people for their communication with the common language people. Due to this there is some difficulties for them to convey the message over long distance. In proposed system the sign language people and common language people will be able to communicate through devices which makes the distance communication possible and easy. In this video call the actions of sign language people is predicted and shown for common language people that they can understand their message directly even on long distance.



REFERENCES

[1] Mitchell, Ross; Young, Travas; Bachleda, Bellamie; Karchmer, Michael (2006). "How Many People Use ASL in the United States?: Why Estimates Need Updating" (PDF). Sign Language Studies (Gallaudet University Press.) 6 (3). ISSN 0302-1475. Retrieved November 27, 2012.

[2] Singha, J. and Das, K. "Hand Gesture Recognition Based on Karhunen-Loeve Transform", Mobile and Embedded Technology International Conference (MECON), January 17-18, 2013, India. 365-371.

[3] D. Aryanie, Y. Heryadi. American Sign Language-Based Finger-spelling Recognition using k-Nearest Neighbors Classifier. 3rd International Conference on Information and Communication Technology (2015) 533-536.

[4] R. Sharma et al. Recognition of Single Handed Sign Language Gestures using Contour Tracing descriptor. Proceedings of the World Congress on Engineering 2013 Vol. II, WCE 2013, July 3 - 5, 2013, London, U.K.

[5] T.Starner and A. Pentland. Real-Time American Sign Language Recognition from Video Using Hidden Markov Models. Computational Imaging and Vision, 9(1); 227-243, 1997.

[6] M. Jeballi et al. Extension of Hidden Markov Model for Recognizing Large Vocabulary of Sign Language. International Journal of Artificial Intelligence & Applications 4(2); 35-42, 2013.

[7] H. Suk et al. Hand gesture recognition based on dynamic Bayesian network framework. Patter Recognition 43 (9); 3059-3072, 2010.

[8] P. Mekala et al. Real-time Sign Language Recognition based on Neural Network Architecture. System Theory (SSST), 2011 IEEE 43rd Southeastern Symposium 14-16 March 2011.

[9] Y.F. Admasu, and K. Raimond, Ethiopian Sign Language Recognition Using Artificial Neural Network. 10th

International Conference on Intelligent Systems Design and Applications, 2010. 995-1000.

[10] J. Atwood, M. Eicholtz, and J. Farrell. American Sign Language Recognition System. Artificial Intelligence and Machine Learning for Engineering Design. Dept. of Mechanical Engineering, Carnegie Mellon University, 2012.

[11] L. Pigou et al. Sign Language Recognition Using Convolutional Neural Networks. European Conference on Computer Vision 6-12 September 2014.

[12] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. http://caffe.berkeleyvision.org/, 2014.

[13] Lifeprint.com. American Sign Language (ASL) Manual Alphabet (fingerspelling) 2007.