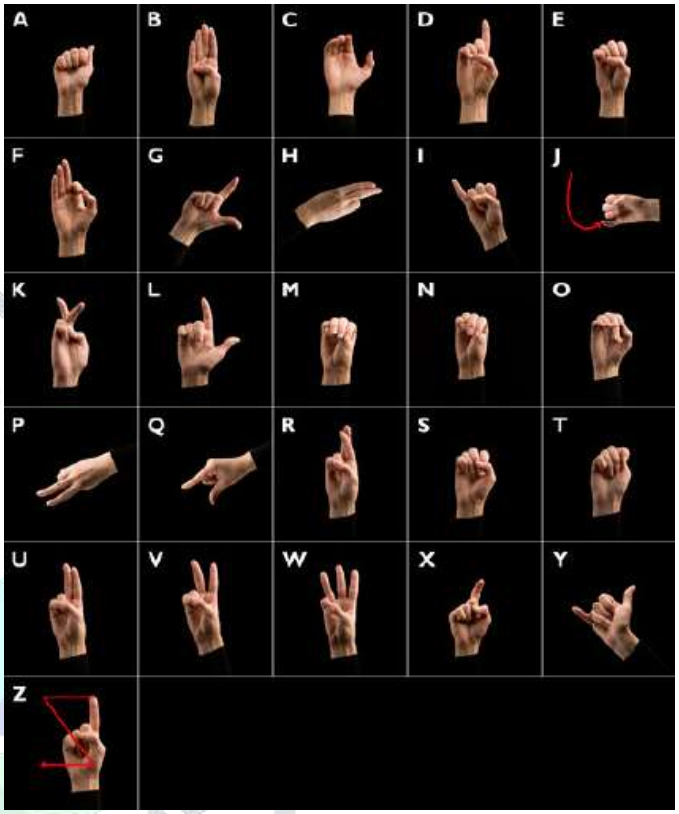
SIGN LANGUAGE ORATOR USING MACHINE LEARNING

SANKET PATIDAR, PRATEEK JANI, ASHOK MAKHIJA

Abstract:
Human beings interact with each other either using different language channel such as words, writing, or by gestures e.g. hand and head gestures, facial expression, lip movements and so on. Sign language is a natural language and as understanding natural language is important, understanding sign language is also worthwhile. Sign language is a communication method for people with hearing and speaking disability. People with hearing and speaking disabilities face problems in communicating with other hearing people without a translator. Even after decades of digitalization some abstract topics such as this one still remain unexplored. In this project, we have proposed an algorithm which is capable of extracting sign's from a live video stream. We are working towards enhancement of deaf people's social life by providing them this sign language orator which would help them to interact with people who are not familiar with sign language. This system can be used for professional conversations as wells as for day to day life chores. This algorithm will help us in converting video of frequently used full sentences gesture into a transcript and then finally convert it into speech. So the signs and their corresponding meanings will be interpreted and identified by using Haar Cascade Classifier. Finally the speech synthesizer will convert displayed text into audio.

I. INTRODUCTION

A machine learning program that detects and recognizes various signs performed by the hand that are been captured and fed the program through a video stream (webcam or a particular IP webcam). This program works on a supervised learning algorithm which is used to create an initial logic tree which helps the algorithm to later on recognize those actions. The machine can recognize and identify only those signs that are fed into it by the user while training the module on a particular dataset which has a finite number of videos (showing signs) and all and all the objects that are stored in the logic tree from which the program derives its answers to recognize the signs and provide you with an audio output.



Sign Language English alphabets representation

Sign gestures can be majorly classified into two types static and dynamic. Static gesture is simpler than dynamic gesture recognition, in static only single image is recognized at a time while on the other hand dynamic gesture is a moving gesture represented by various images. Sign language is used by people who have speaking or listening disability so that they can exchange information between other people and within their own community. This can be used on wide scale to make most of the public understand what a person is trying to convey to the world through sign language. This will prevent the disabled people for hiring other people who can speak and understand sign language just to be their communicator.

II. LITERATURE SURVEY

According to the World Health Organization (WHO), around 5% of world population belongs to the people with the hearing and speaking disability that totals

over 360 million people across the globe. The majority of these people live in countries with comparatively low incomes. Sign Language is an independent language which is different from spoken/written language, the basic difference is it has limited vocabulary compared to written/spoken. Sign language is not the same in every country, different sign languages are used in different countries or regions. As you may know there are two separate languages ASL and BSL among which ASL is the most widely used signed language.

The real time input video is initially segmented into frames and the systems that are developed works on images to recognize sign language, the main device used is camera, which helps to capture a gesture image which is an input data for Sign Language Recognition (SLR) and this image is then processed and the result is obtained. There is also other device like Microsoft Kinect which is used to capture image. It is widely used by researchers because of it features like color video stream and depth video stream.

A general framework for hand movement detection and analysis involves stages such as motion detection with the help of background modelling and foreground segmentation, object classification, motion tracking and activity recognition.

This paper describes a technology in which real time videos are analyzed and are used for hand movement detection and recognition, thus helping them to convey what they want to explain or tell, in the form of transcripts and converts it into audio. The system developed identifies sign language (non-verbal communication) done by using some hand gestures and the machine is trained to recognize some daily frequently used gestures and convert them into verbal communication so that the other people who are not familiar with this non-verbal communication can understand easily which will ultimately prevent them to hire a person to communicate for them and save their money.

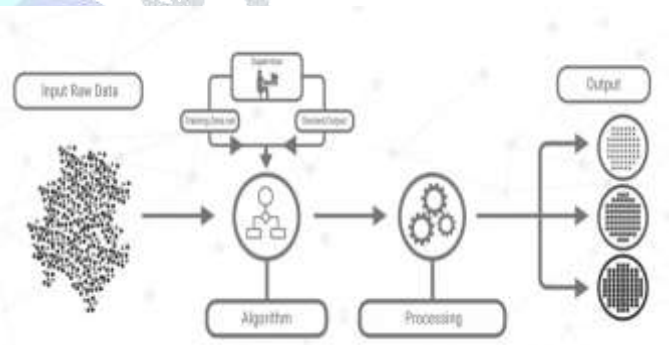
A Phoneme based method is also used to recognize the sign language. There are 44 phonemes in English language; therefore 44 gestures can be formed. There are 11 categories of gestures which are formed in ASL. The right hand shows categories, left hand shows sign in category. For prepossessing and filtering the image, RGB color space is used. RGB algorithm shows the head and hands region. The next step consists of using the vertical interleaving method for image compression. The features of the image are extracted using the 2-D moment invariant which are then fed to the neural network which recognizes the equivalent text.

III. MACHINE LEARNING

Artificial intelligence (AI) is a subarea of computer science that emphasizes the creation of automated machines that work and reacts like humans. Machine learning system is a branch of artificial intelligence based on an idea that a system can produce general hypothesis by learning from data provided, identify patterns and make decisions with minimal human intervention. Machine learning is important because as models are exposed to new data, they are able to create a predictive model capable of inferring annotations for future data. They learn from previous computations to produce reliable, repeatable decisions and results without needing multiple manual edits to the program.

Supervised Learning

Supervised Learning is a computational task of automated data (the training data set) to produce general hypothesis. The training data set comprises of input and output pairs which are used to train the model and determine a hidden pattern. This hidden pattern is then used for recognition of specific patterns when working data is provided in future.



Supervised learning is in which you have input variable (a) and an output variable (b), and you apply an algorithm to learn the mapping function from the input to the output.

$$Y = f(A)$$

The aim is to approximate the mapping function so well that when you have new input data (a), you can predict the output variables (B) for that data. Supervised learning is called so because the method of algorithm learning from the training dataset can be thought of as a trainer supervising the learning process. We know the accurate results; the algorithm iteratively makes predictions on the training data and is corrected by the trainer. Learning stops when the algorithm achieves an agreeable level of performance.

SVM (Support Vector Machine)

It is a supervised learning model in machine learning, algorithm that analyses data for classification and regression analysis. Mainly used in text categorization, image classification and handwriting recognition. It sorts data into two categories, place the trained data into one of them and the other one belongs to new data.

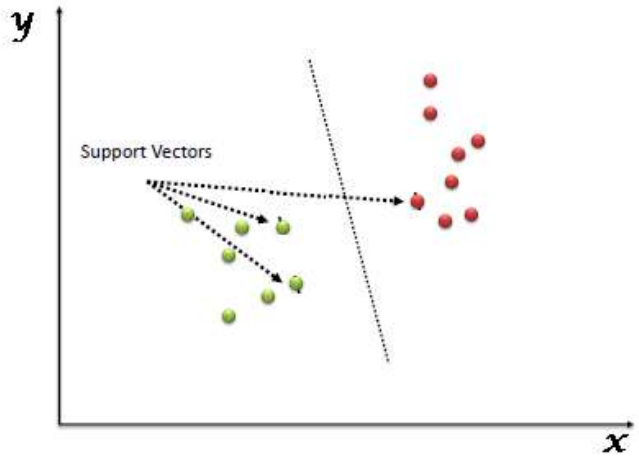


Fig: Support Vector Machine Graph

The two observations that we obtained are now connected through vectors as shown above in the graph and the inner product between two vectors is the sum of the multiplication of each pair of input values.

For example, the inner product of the vectors [4, 5] and [6, 7] is 4*6 + 5*7 or 28.

Now the equation to make prediction for a new input by using the dot product between the input (x) and each support vector is calculated as follows:

F(x) = B0 + sum (ai * (x, xi))

The equation calculated contains the inner product of a new input vector (x) with all support vectors in trained data. Coefficients B0 and ai (for every input) evaluated from the trained data by the learning algorithm.

Deep Neural Networks

Neural networks is a set of algorithm which is used to recognize the relationship between the underlying set of data in way that a human brain works which constantly try to recognize patterns and categorize and classify information. It is an excellent tool for finding patterns which are complex as well as numerous for programmers to teach the machine to recognize. It usually involves a large number of processors that are operating in parallel and are arranged in tiers. Hidden layer is in between the two input and output layers.

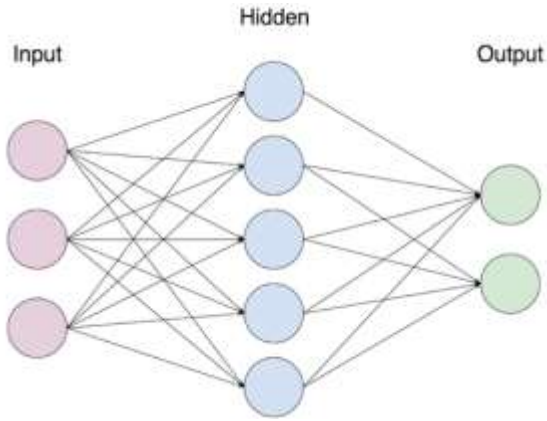


Fig: Single hidden layer neural network

Deep neural network is similar to deep learning, with certain level of complexity and has more than two layers. The different layers of such a system could be seen as a nested hierarchy of related concepts or decision trees. Deep neural network systems need large quantities of data in order to be trained as the systems learns from exposure to huge number of data points. Google Brain learning to recognize cats after being shown over ten million images can be seen as an early example of this.

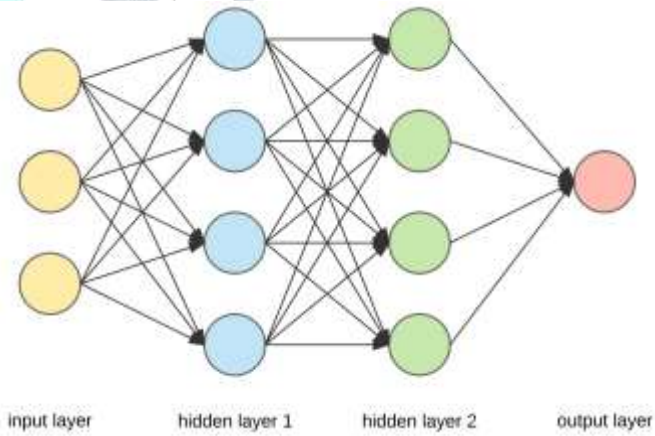


Fig: Multi- layered Deep Neural Network

OpenCV

OpenCV (Open Source Computer Vision Library) is a library developed by Intel of programming functions and is released under BSD licence hence it's free for both academic and commercial use. It is a powerful library designed to work on real time applications with a strong real-time efficiency. It is written in C++ and it's primay interface is based on C++ this makes OpenCV portable to almost any commercial system includesPython, JAVA, MATLAB/OCTAVE interfaces which are supported by Windows, Mac OS, iOS, Linux, FreeBSD, OpenBSD and Android as it was designed to be a cross-platform, the library can take advantage on multi level processing thus makes it easy for businesses to utilize and modify the code.

IV. PROPOSED METHOD

In this paper we aim to enhance the already existing supervised learning model by using the Support vector machine (SVM) model to get better organized results. The SVM technique is already a successfully proved technique to recognize images but the main problem with it is the results are not processed in real-time and sometimes it results in false recognition. We focus on trying to improve these limitations by combining the SVM unsupervised learning model with a supervised learning neural network model which will process and recognize the signs in real-time and improve the performance and will improvise the learning and recognizing capability of the system when new data is processed.

Initially, the user of this system will enter his details. Since this system uses a video based approach, the user will record a video in ISL using a camera provided by the system or any other hardware such as webcams. The application will then convert the video into a sentence in simple English. This system will enable him to communicate with others without the need of a human interpreter. Researching all the above techniques to carry out the various stages of the project, we have proposed the techniques which we will use in our solution.

- **Frame extraction**

The video which we have recorded will be saved in a pre-determined folder. The video will be taken from here and frames will be extracted. An interval is defined to extract the required frames.

- **Segmentation**

This step attempts to simplify an image by partitioning it into multiple segments of sets of pixels. We will use the CIE Lab color model because it carries out fast segmentation and it is an absolute color space so it defines the colors exactly without depending on input devices. It also includes more colors than other color spaces.

- **Feature Extraction**

This step will extract the important features from the images to feed them to the neural network. Once the image has been cleaned and segmented, important features need to be extracted from it to feed into the network for them to be classified. It is needed to reduce the amount of memory and speed required to classify the data.

- **Neural Network**

This step uses the extracted features as an input. We are using the Error Backpropagation Training Algorithm to train the network. This is because it will take a lot of training for the network to classify the images into the correct signs and the errors in each step have to be continuously reduced to get a good success. The errors are propagated backwards from the classifier so that the network can learn. Our output is expected to be a sentence in English, in which the words correspond to the interpreted signs classified by the machine.

V. CONCLUSION

This is to conclude that the project that we undertook was worked upon with a sincere effort. The project can be completed with the current available technology and we hope that resources and minds would work on this to convert this paper into a viable prototype. The prototype would be of immense assistance for those who are deaf or have speaking disability.

VI. REFERENCES

[1] Priyanka Mekala, Ying Gao, Jeffrey Fan, Asad Davari "Real-time Sign Language Recognition based on Neural Network Architecture"April 2011 International Journal of Information and Communication Technology

[2] Suharjitoa*, Ricky Andersonb , Fanny Wiryanab , Meita Chandra Ariestab , Gede Putra Kusumaa "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output" Suharjitoa*, Ricky Andersonb , Fanny Wiryanab , Meita Chandra Ariestab. ScienceDirect Procedia Computer Science 116 (2017) 441–448

[3] Tharadevi. R.V "Hand Gesture Recognition using Sign Language" Imperial Journal of Interdisciplinary Research (IJIR) Vol-3, Issue-9, 2017 ISSN: 2454-1362

[4] Fabio Fabris, Joa˜o Pedro de Magalha˜es, Alex A. Freitas. "A review of supervised machine learning research" 28 October 2016 / Accepted: 21 February 2017.

[5] Osisanwo F.Y., Akinsola J.E.T., Awodele O., Hinmikaiye J. O., Olakanmi O.,Akinjobi J.

"Supervised Machine Learning Algorithms: Classification and Comparison" International Journal of Computer Trends and Technology (IJCTT) – Volume 48 Number 3 June 2017.

[6] Rashmi B. Hiremath, Ramesh M. Kagalkar "Review Paper on Sign Language Recognition Techniques". International Journal of Computer Applications (0975 – 8887) National Conference on Advances in Computing (NCAC 2015) .

[7] Esam Kamal MARIED, Mansour Abdalla ELDALI ,Osama Omar ZIADA," A Literature Survey of Deep learning and its application in Digital Image Processing", UNIVERSITY OF TURKISH AERONAUTICAL ASSOCIATION DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING : (EEE 573),available : https://www.researchgate.net/publication/317346042_A_Literature_Study_of_Deep_learning_and_its_application_in_Digital_Image_Processing

[8] Akshay Jadhav, Gayatri Tatkar, Gauri Hanwate, Rutwik Patwardhan. "Sign Language Recognition" International Journal of Advanced Research in Computer Science and Software Engineering.

[9] Ms. Sonali. B. Maind, Ms. Priyanka Wankar "Research Paper on Basic of Artificial Neural Network" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 1.

[10] OpenCV Documentation https://opencv.org/about.html

[11] Theodoros Evgeniou and Massimiliano Pontil "SUPPORT VECTOR MACHINES: THEORY AND APPLICATIONS". https://www.researchgate.net/publication/221621494_Support_Vector_Machines_Theory_and_Applications.