Sign Language Prediction using Machine Learning

Vivek Bafana[1], Omkar Dukhande[2], Rudra Palande[3], Priya Porwal[4]

[1] *(Department of Computer Engineering, Atharva College Of Engineering, University of Mumbai, India*
Email: bafna.vivek165@gmail.com)
[2] *(Department of Computer Engineering, Atharva College Of Engineering, University of Mumbai, India*
Email: omkardukhande19@gmail.com)
[3] *(Department of Computer Engineering, Atharva College Of Engineering, University of Mumbai, India*
Email: rudrap199727@gmail.com)
[4] *(Department of Computer Engineering, Atharva College Of Engineering, University of Mumbai, India*
Email: priya.porwal20@gmail.com)

Abstract: *A real time sign language is important in facilitating communication between deaf people and general people. Machine learning relies on recognition of a pattern in data and the concept that computers can learn without being programmed by programmers to perform certain specific tasks; researchers in artificial intelligence wanted to see if it was possible to teach computers to learn from data and to improve upon it. Using Neural Network and Machine Learning algorithms application can convert the sign language and also help in predicting the next letter which we make use of daily. This paper is a brief review of a machine learning application which is Sign Language Prediction.*
Keywords: *ANN (Artificial neural network), Database management system, Image recognition, Machine learning, Neural networks, Prediction Algorithm, Reinforcement learning Applications, Supervised and Unsupervised learning.*

I.   Introduction

Sign language is a communication system using visual gestures and signs to communicate, as used by the deaf people. These visual gestures and signs always convey meaning. This proposed system takes the input from the camera which captures gesture and posture using OpenCV and conveys the message of the input and then using prediction algorithm which tends to complete the message even when the gesture or the posture has not completed conveying. In a survey, it is found as most of the people in India are unaware of sign language. Deaf people face problem talking to normal people and normal people face understanding the deaf people. There are applications which help in conveying the meaning of sign language and our application is the same but also has an addition of prediction methodology which saves the time and also ease communication between the deaf and normal people.

II. Machine Learning

Machine learning is one of the field in computer science which is developing and has a vast application in many fields. Arthur Samuel coined the term 'Machine learning' as a field of study that gives computers the ability to learn without being programmed by programmer explicitly. A more formal definition based on experience, task and its performance measure was given by Tom Mitchell as a computer program which is said to learn from experience (E) according to some task (T) and some performance measure (P), if its performance on T, as measured by P, increases with experience E then that program is called a machine learning program.
Machine learning can be broadly divided into four types: Associative learning, Supervised learning, Unsupervised learning and Reinforcement learning. Supervised learning process is based on the comparison of actual output and desired output.

Unsupervised learning can be explained as learned by its own and discovering and adopting accordingly, based on the input pattern. It is sometimes called as clustering technique or clustering algorithm. Reinforcement learning is similar to supervised learning. But, in some case, less information might be available. Association learning involves rule-based machine learning method used for discovering interesting relations between variables in large databases. Machine learning has a vast application in fields like Business, data analysis, Robotics, Military and thus list goes on.

III. Related Work

This particular field of sign language prediction is been improved for almost 20 years along with speech recognition. Analyzing and processing of audio signal is easier as they being one-dimensional whereas

image being two-dimensional, it is trickier to analyze and predict the output from images. The images should not be noisy and they are to be converted into proper format which are easier to analyze but does not provide wrong result. Here we implement methods present in reference [6] for grey scaling and reference [8] for noise removal and filtering and [4] for hand posture estimation. In order to convert image to HSV and to apply masks the method shown in reference [7] can be used.

The database used contains images which may be grey-scaled or HSV and the corresponding meaning of those images. The research for these databases and its accuracy is done in reference [9]. The idea of tracing the hand contours by Zernike moments is shown in [3] which can be implemented; it provides accuracy but at the cos of speed of processing. The methods used in [1] and [2] are used to predict one letter at a time and they do not store the previous outputs for future use. The system in this paper uses some of the techniques used in the given reference and improves upon it by adding some more phases in the system which improves the application and scope of the system in general.

IV. Literature Review

Authors: Maher Jebali, Patrice Dalle, Mohamed Jemni [1]
Their proposed system uses a modeling approach for sign language recognition which is introduced through the dialogue recognition between deaf people and the signing avatar. This approach is only a method of converting sign gestures to their equivalent meaning. As this type of approach favors in field of Robotics, Artificial Intelligence, etc. This Human Computer Interaction (HCI) is been going on from a long time. This system mainly focuses on two elements they are context and the prediction. The context deals with the sign language which is input taken from cameras from the deaf people and then the prediction element to convey the message of the corresponding sign gesture.

Authors: Ashish S. Nikam, Aarti G. Ambekar [2]
In this system they have used camera instead of sensors for capturing images to process it further. Next step is Image Processing which contains removal of unwanted noise, adjusting brightness and contrast of the image. In this process contains image enhancement, segmentation and color filtering process. Noise present in the image is removed by using morphological operations such as Erosion and Dilation. They have used thresholding process to convert the image into binary format. Hull algorithm can detect finger point and the number. Also it contains preliminary handling of the images so the noise decreases of the input and recognition can be done by matching the templates in the dataset

Authors: Anup Kumar, Karun Thankachan and Mevin M. Domini [3]
This proposed system uses classification model. This system classifies the input gesture using OpenCV library and then removing the noise from the input only the sign conveying image. Then image is converted and zernike moments and curve features are used for the feature extraction. Support Vector Machine algorithm in python is used as the classification model. The input is converted into frames and then smoothened. The zernike moments from the frames helps in the extraction of the images and then compared with database to find the equivalent meaning of the gesture.

Authors: Jeevan Musale, A P Mane [4]
This paper proposes a programmed gesture recognition approach to Indian communication via gestures. Their system uses live images of sign to predict the sign using Support vector machine. The system first converts the image to be process able by the help of HIS shades and then the SVM does the prediction part. Hand gesture recognition framework comprises of the accompanying steps, picture procurement, Pre-processing, hand segmentation, and hand identification. This system demonstrates steady and constant execution for progressed transforming speed. That algorithm makes utilization of classifier SVM for proper recognition with a low rate of error.

Authors: Y.Fang, K.Wang, J.Cheng&H.Lu [5]
In this paper, they proposed a robust real-time hand gesture recognition method. In this method, firstly, a specific gesture is required to trigger the hand detection followed by tracking; then hand is segmented using motion and color cues; finally, to order to break the backdraw of aspect ratio encountered in learning hand based gesture methods, the space scale feature detection is combined with gesture recognition. Applying this method to navigation of image browsing, experimental results showed that this method achieved good performance. Compared with other interaction approaches, such as keyboard, pen, mouse etc., vision based hand interaction is more efficient and natural.

T.Dinh, V.Dang&D.Duong [6]
In this paper, they propose a hand gesture classification system able to efficiently recognize 24 basic signs of American Sign Language. In this system, computational performance is achieved through the use of a boosted cascade of classifiers that are trained by AdaBoost and informative Haar wavelet features. A new type of feature to adapt to complex representation of hand gesture is also proposed inn this paper. Experimental results showed that the proposed approach is promising. Hand gesture recognition is a vital component in applications such as human computer interaction, robot control, and disable assistance systems in which robustness and performance are the primary requirements.

Authors: Kabil Jaballah and Mohamed Jemni [7]
This proposed system deals with the 3D Sign language data that is generated through dedicated signing avatars system. It does this using 3D devices like data-gloves and methods H-anim compliant software. Also estimators are used which are database predefined and the descriptive classifiers. Using these estimators the compared with each other and finds the joint angle which can be referred as similar points in the avatars and the degree of freedom is present which makes easy to find the joint angles. Now the designed rules are applied to generate the value of the avatar.

Authors: T Starner, A Pentland [8]
This Proposed system uses more than just finger singing as many signs conveys message which are not finger singing. This model is build with the database which focuses on such types of avatars that is 3D models. This uses HIdden Markov Models. Here hand tracking stage does not tend to find description of the avatar. The tracking process produces only a course description of the hand shape, orientation and trajectory. The skin color is used to identify the joint angles to classify and remove the noise and using classification they can be converted to the resemblance with the database and convey the meaning of the sign or avatar.

Authors: D Deora, N Bajaj [9]
This paper represents a framework for a Human Computer interface which is capable of recognizing the gestures which are present in the Indian Sign Language. The Indian sign language is complex due to involvement of both the hands and also the overlapping of the hands. Also some signs involve motions in them so recognition becomes difficult sometimes. This system can recognize alphabets and numbers. They are trying to extend the system to recognize words and sentences. Recognition is performed with PCA and also with neural networks.

Report on present investigation
- About millions of people across the world (Over 5% of world's population) have to live with disabling hearing loss.
- A large percentage of those live in low and medium income countries which does not have access to proper ear and hearing care services.
- Among those people, not everyone can afford devices that aid in hearing or any cochlear implants.
- The use of a sign language system can help them to communicate among themselves as well as with other people.
- Research is ongoing on communicating with Robots used in industries and factories using sign language.

V. Proposed System

Sign language is a communication system using visual gestures and signs to communicate, as used by the deaf people. These visual gestures and signs always convey meaning. Over 5% of world population have to deal with hearing problem. Thus a real time sign language is important in facilitating communication between deaf people and general people. Machine learning relies on recognition of a pattern in data and the concept that computers can learn without being programmed by programmers to perform certain specific tasks. This system should be able to provide proper communication between them and a user friendly interface is to be provided for the communication.

5.1 Problem Statement
Deaf and dumb people have problem in communication with another people who has less knowledge of sign language. Thus deaf people loss interest in common activities and avoid communication with normal people and live in isolation. Sign language system overcomes this problem by predicting the text output.

5.2 Solution

The system consists of an image processing system that captures and filter the image or posture, and extract the required data that the prediction system can utilize to provide an optimal output; a machine learning model implemented using ANN or a Classifier that takes the data from the image processing system and analyzes it with the available system database to find for a match and thus provide a letter or a word to the pre-prediction system; Database of sign language which contains images of sign along with its meaning or associated word which describes it in that sign language; and the pre-prediction system that provides a complete word or sentence based on the letters that the model provides, frequency of the word used and the recent history it gives a set of proper, optimal output of words and sentences from which the user can select ; it can save time and make system more effective and easy to use. The complete process flow of the system is shown in fig.3.

5.3 Image Processing System

The first phase of the system is the image processing system. This part of the system can be implemented using OpenCV lib., which is compatible with python programming language which is used to implement majority of the system. The main task of this phase is to capture the user's gestures and sign and to prepare it to for the next phase.

In this system Firstly the image is flipped and then we may use HSV for segmentation and Zernike moments and curve features for feature extraction. The image can be converted to HSV because HSV colors are much easier to work on and isolate color then RGB values.

To isolate the color required we can apply multiple threshold mask which will set the required pixel which matches the criteria of hand boundary as 1 and the rest as 0 to get an isolated image with only boundaries of hand highlighted while rest is black. This method is known as greyscale. This method will not work best when the image is noisy. Noise may occur in image if the camera used to capture the image uses analog-to-digital convertor which may introduce noise in the system. A solution to this is to use a gaussian blur before applying this particular method. The image can be converted to grayscale to improve the extraction process and also the background noise and errors are removed in this stage. The data extracted is then passed to the next phase of the system.

5.4 Machine Learning System

The data extracted from the image processing system is then provided to the prediction model. This model can be implemented using neural networks that takes the data from the image processing system and analyzes it with the available system database to find for a match and thus provide a letter or a word to the pre-prediction system. Basic machine learning algorithms like K-nearest algorithm, Random forest classifier or Haar classifier can be further implemented to ease the prediction process and to provide optimal output based on the accuracy of the implemented model. The image extracted is linearized i.e. if the image pixel matrix is of size (n,m), it is converted to (1,n*m) for proper comparison of image with the database image. The process flow of machine learning phase is shown in fig.4.

The most effective method in this case is a Haar cascade classifier. Haar-like features are digital image features which are used in object recognition in an image. They derieve their name to their similarity with Haar wavelets and were used in the first real-time face detector. A Haar cascade classifier is a classifier which is used to detect the things for which it is trained for from provided input. The Haar Cascade classifier does this by superimposing the positive image over a set of negative images. The training can be done on a server and in various stages. Better results are obtained by using images with higher quality and increasing the amount of stages for classifier's training.
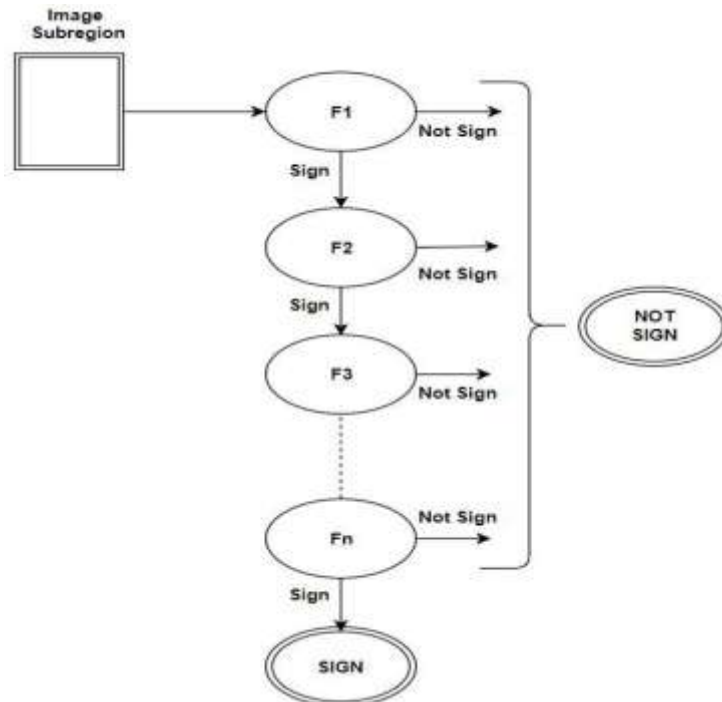
Fig 1. Working of classifier

5.5 Language System

The prediction model uses a database of sign language which contains images of sign along with its meaning or associated word which describes it in that sign language. The images in database are stored as SQL BLOB (Binary large object). The images in database are already stored in HSV and proper mask are applied for faster comparisons. The system can be made such that the database can be further expanded and trained to model in the future.
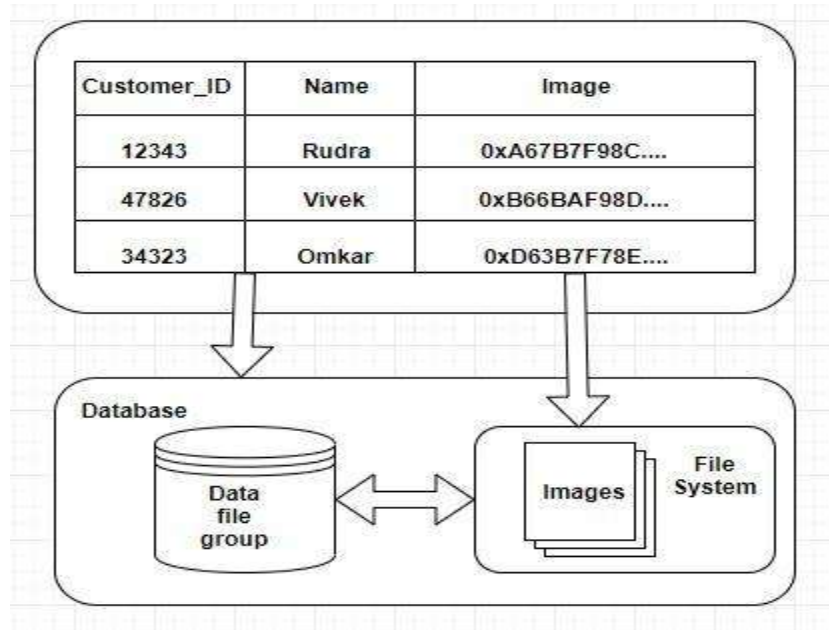

Fig 2. Representation of database

5.6 Pre-Prediction System

Pre-prediction system that provides a complete word or sentence based on the letters that the model provides and based on the frequency of the world used and the recent history it gives a set of proper, optimal output of words and sentences which the user can select; it can save time and make system more effective and easy to use.

This phase uses a separate word database which stores the frequently used words along with their frequency and last use. This database can be either a csv file or sql file. The system working can be explained with example: Suppose the system scans a sign which the model predicts as 'H'. it then stores it in a buffer and then scans for next sign. Suppose the next sign turns out to be 'E' ; then the pre-prediction model can give all the words from the word database that starts with 'HE' which are frequently used or recently used like 'HE','HELL','HELLO', Etc. Thus the user can select from those words or he can provide more letters to get more proper word. This method can further be applied to statements with proper algorithm which can further increase the accuracy and potential of the system. Finally this phase provides with either a letter or word. The user may continue to add more letter or word to the previous input and the previous input will be stored in a buffer which may be updated. The user may continue this till the word or sentence he wants is obtained.
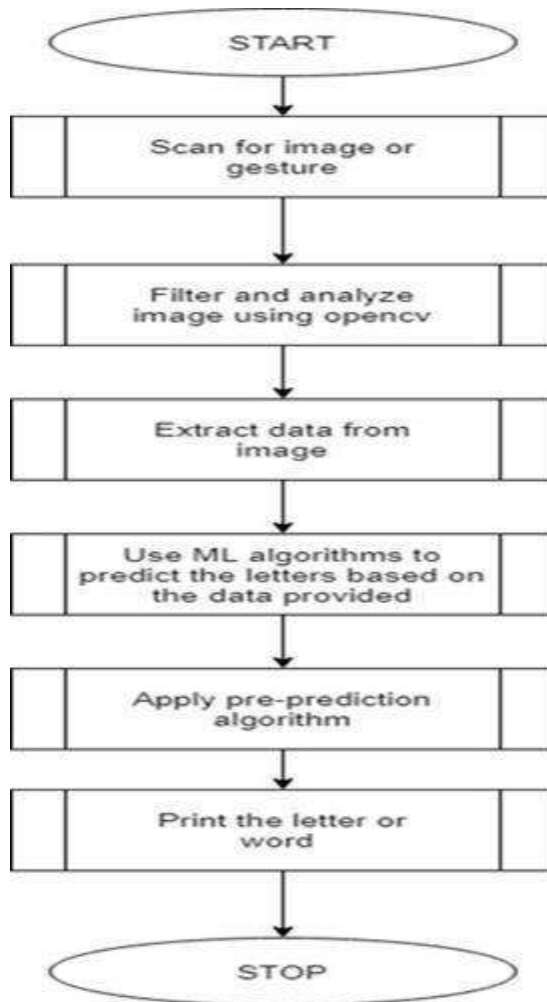

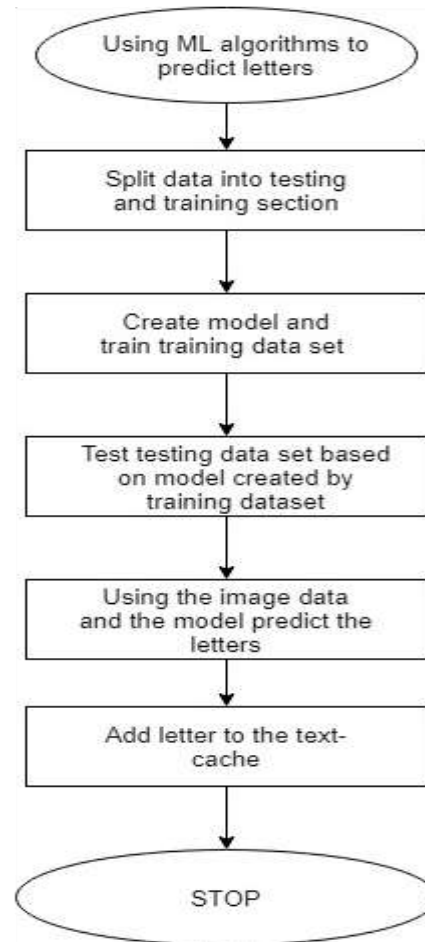**Fig 3** Process flow chart of system   **Fig 4** Process flow chart of the machine learning system

VI. Conclusion

Sign language prediction system can help deaf people to have easy communication with computer. The suggested prediction based recognition framework analyzed temporal visual signal obtained from the cameras which is changing dynamically in each level of dialogue, showing a system that provides results in real time which can be used by the deaf with no exception. Database can be increased corresponding to various languages which lead to the prediction of different languages. It will be capable of predicting the output in different languages. It can be deployed in applications on various platforms. Depends on database it can achieve maximum accuracy.

References

[1]   Maher Jebali, Patrice Dalle, Mohamed Jemni.: Towards Sign language recognition system in Human-Computer interactions. *Information and Communication Technology and Accessibility (ICTA), 2013 Fourth International Conference on. IEEE, 2013*.

[2]   Ashish S. Nikam, Aarti G. Ambekar.: Sign Language Recognition Using Image Based Hand Gesture Recognition Techniques. *2016 Online International Conference on Green Engineering and Technologies (IC-GET)*

[3]   Anup Kumar, Karun Thankachan and Mevin M. Domini.: Sign Language Recognition. *Recent Advances in Information Technology (RAIT), 2016 3rd International Conference on. IEEE, 2016*

[4]   Jeevan Musale, A P Mane.: Indian sign language recognition system using Combinational features. *In: GRD Journals- Global Research and Development Journal for Engineering, Volume 1, Issue 4, March 2016*

[5]   Y.Fang, K.Wang, J.Cheng&H.Lu.: A real time hand gesture recognition method. *IEEE, 2007.*

[6]   T.Dinh, V.Dang&D.Duong.: Hand gesture classification using boosted cascade of classifiers. IEEE 2006. *University of Waterloo, Fall 2014.*

[7]   Kabil Jaballah and Mohamed Jemni .: Hand Location Classification from 3D Signing Virtual Avatars Using Neural Networks. *ICCHP (2) 2014: 439-445*

[8]   T Starner, A Pentland .: Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on pattern analysis and machine, 1998*

[9]   D Deora, N Bajaj: Indian Sign language recognition. *2012 1st International Conference on Emerging Technology Trends in Electronics, Communication & Networking*