ANDROID BASED SIGN LANGUAGE RECOGNITION AND TRANSLATION

[1]Mrs. Sneha N P, [2]Nabeela Akram, [3]Rajeshwari S, [4]Rakshitha L, [5]Trishul S

[1]Assistant professor, [2]Research Scholar, [3]Research Scholar, [4]Research Scholar, [5]Research Scholar
[1]Department of Computer science and engineering,
[1]ATME College of Engineering, Mysore, India.

Abstract: The aim of this paper is to outline an advantageous framework that is useful for the general population who has speech and hearing disability and those who utilize exceptionally basic and powerful communication strategy through gestures. It can be utilized for changing over gesture based communication to voice based communication in regional language which includes English and Kannada. It captures the signs and directs on the screen as composing. Inability to speak is considered to be true disability of these people. People with this disability use different types of sign languages in order to communicate with others, there are n number of languages available for their communication one such common language of communication is Indian sign language. It allows people to communicate with human body language where the each word has a set of human actions representing a particular expression. The motive of this paper is to convert the human sign language to voice.

Keywords - Regional Language, Gestures, Composing, Indian Sign Language (ISL).

I. INTRODUCTION

The aim of this paper is to improve the communication with the people who has speech and hearing difficulties and sing any sign language in order to express them. There are several sign languages available such as American, British, German, French, Italian, Turkish and Indian Sign Languages. One of the sign language known as the American Sign Language is well-known sign language. It is the best studied sign language in the world. The grammar of American Sign Language has been applied to other sign languages especially in the British Sign Language. It is to design a convenient system that is helpful for the people who have speech and hearing difficulties and in general who makes use of very simple and effective method sign language. This system can be used for converting sign language to voice in regional language and also voice to sign language. The major goal in this project is that we develop a technology based on vision for recognizing the gestures and translating continuous sign language to the text.

A motion capture system is used for sign language conversion and a voice recognition system for voice conversion. Where the motion capture is sometimes also referred to as mocap, which is the process of recording the movement of objects or it can be the movement of people. It is used in the fields such as in entertainment shows, sports, and medical applications, and also for validation of computer vision and in robotics. It captures the signs and dictates on the screen as writing. It is very much easier to find a wide variety of sign languages all over the world and almost every spoken language has its respective sign language. So, there are about more than 200 languages available. Here it presents a Android mobile interactive application for automatic translation of Indian sign Language into speech or the text in English and Kannada to assist the hearing and/or speech disability people to communicate with the people having the hearing difficulty**.** This developed sign language translator must be able to translate alphabets from (A-Z) and numbers from (0-9). Many languages are being spoken all around and in the world. So this system aims at giving the voice as output in various regional languages which includes English and Kannada. This project can be modified to make it compatible with mobile phones. We can increase the range of product by using more powerful trans-receiver module. In this project the disabled person must have a Smartphone and he can install the application in his or her phone by downloading it from the play store once it is hosted. After that user can register with the credentials required by the application and then start to capture image so, that the sign gestures of hand can be recognized and the output can be heard through voice using the application.

1.1 Advantages

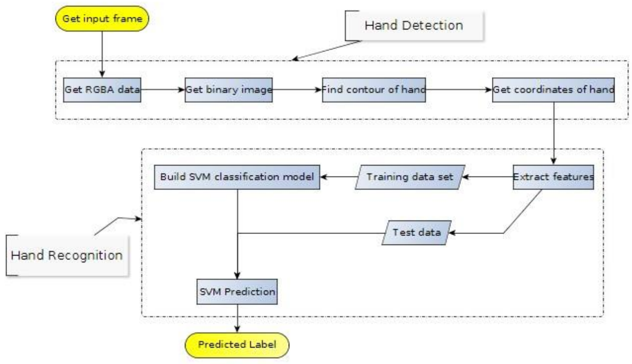
- The Person with speech and hearing disability need not be educated or know English Language in order to communicate.
- The opposite person just needs to listen to the output that is being given by the application in order to understand what the person with disability is trying to convey.
- The Person need not learn to write in order to communicate; instead he must know the basic regional language or the English language.

II. LITERATURE SURVEY

Sign language recognition and translation system mainly has two very well known approaches i.e., Image processing technique and another one is microcontroller technique and sensor based data glove technique. Image Processing is simple understood as image + processing, where the frames are captured from the image. The Data glove technique is the one where with the help of sensors it senses the gesture made and displays the text in English as the output in the application. These approaches are also known as vision based and sensor based techniques. In the image processing technique camera is used to capture the image or a video, in the images that doesn‘t undergo any changes i.e., the static images are analyzed and image recognition is using carried out using algorithms that outputs sentences in the display. The algorithms that are used in vision based sign language recognition system are HMM, ANN, and SAD. Where HMM is abbreviated as Hidden Markov Mode, and ANN as Artificial Neural Networks and SAD as Sum of Absolute Difference. The motivation was to create a object tracking application system that was used to interact with the computer and develop an virtual human computer interaction device. The project used webcam to recognize the gesture positions made by hand using contour recognition technique and outputs the audio as an output onto the personal computer screen, which was helpful for the normal people to understand what exactly the opposite person with speech and hearing disability is trying to convey.Here the speech and hearing disability people need sign language to communicate with each other and with other speech and hearing disability people. Moreover, the several ethnic groups that make use of completely different phonologies. (Examples include Plain Indians Sign Language and Plateau Sign Language) i.e., they have used sign languages to communicate with other ethnic groups. But many common people are unaware of these sign languages and it will be a great difficulty for them to communicate with speech and hearing disability people. Some of the disadvantages here include that the people must have their Personal Computers with them in order to communicate with normal people. One must have to be educated in order to understand what is being conveyed through English writing. Children with hearing difficulties analyze their English writing and makes tailored lessons or videos and recommendations. This paper is a small step towards helping a physically challenged people and lot more can be done to make the product more sophisticated, user friendly and efficient. Using more memory and powerful microprocessor, more languages can be covered.

III. METHODOLOGY

The system architecture here represents the design phases which are divided into five modules those are User Registration and Login, Capture frames, Hand Detection, Hand Recognition, and Text to speech translation.



System Architecture

1. *User registration and login:*

The user registration page is created, where it asks the user to enter the details like email ID and password where it will be checking for the correct pattern of email. The entered details are stored in Firebase. The entered email is matched with the format and if it is correct then it simply takes to password field otherwise, it asks the user to enter the email in a valid format. Similarly, it asks the user to enter the password in that format if any criterion is set. Then, the entered details are stored into the Firebase. Firebase is nothing but a backend platform for building Web, Android and iphone Operating System (IOS) applications. It offers the real time database. Features of firebase are authentication, storage, database, etc. Authentication feature is used for user registration and login. If it is a new user then it takes the user to the registration page. If the user's data already exists in the database then it takes it to the login page.

2. Capture frames:

It is the second module in the design phase where once after the registration process in finished it takes the user to this page. When the android application is clicked and the front camera is opened from which it tries to take up the video, where the camera present starts to capture the frames. The input frames are captured for every millisecond (ms) and the each frame should contain at least minimum of 100 pixels to capture the gestures. The captured gestures are then sent to OpenCV where the further processing would be taking place.

2.1 Image Preprocessing

The first stage of the content identification is image pre-processing which is done with the help of the cropping, clipping and other process. Before processing the image need to be converting into the grayscale image because it provides the better results when compared to the color image processing.

2.2 Image Segmentation

The next stage is image segmentation which is the process of partitions an image. It's the process of segment the similar attributes into image which is done with the help of the KNN classification approach.

3. Hand detection:

It is the third module in the design phase where once after the capture frames module is done this module will be in process where the hand detection or the gesture recognition is done. First it tries gets the input frame and then it finds the RGB input (which means the Red, Blue, and Green) and then it tries to find the contour edges of hand and also the coordinates. Contours can be explained simply as a curve joining all the continuous points (along the boundary), having same color or intensity. The contour is a useful tool for shape analysis and object detection and recognition Since OpenCV 3.2, findContours () function no longer modifies the source image but returns a modified image as the first of three return parameters. In OpenCV, finding contours is like finding white object from a black background. So remember, that the object to be found should be white and the background should be black. To draw the contours, cv.drawContours () function is used. It can also be used to draw any shape provided you have its boundary points. After finding these things it is converted into binary image which is done by using OpenCV library functions.

OpenCV: It is abbreviated as *Open Source Computer Vision***.** It is a library of programming functions mainly aimed at providing a real time computer vision. It has C++ (plus plus), Python and Java interfaces and also it supports Windows, Linux, Mac OS (Media Access Control), iOS and Android. The NDK is abbreviated as the *Native Development Kit* which is a set of tools that allows one to use C and C++ code or the programs with Android. 'Emgu CV' is a cross platform and also it acts as a wrapper to the OpenCV image processing library. For Example: Linux is a core component on which the fedora and ubuntu acts as a wrapper to it.

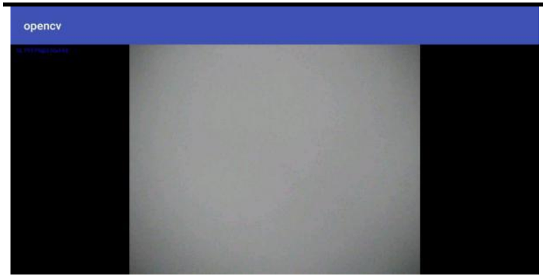
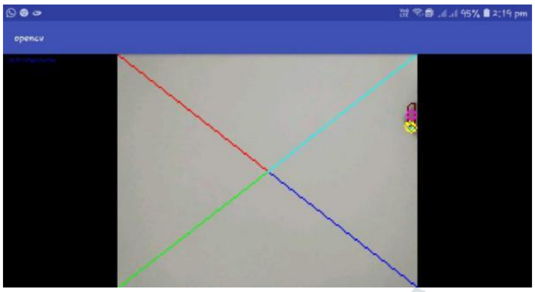


Image - when camera is opened.



The coordinates of a line.

The Contour edges of the gestures can be identified using RGB Color model 'RGB Input Frame' where the each pixel stores three values where R means red and varies from 0 to 255. Similarly, G and B mean green and blue representing values varying from 0 to 255. Each number between 0-255 corresponds to a shade of corresponding color. The Depth (which represents maximum number of bits) of a RGB image is 8 bits and the number of channels for a RGB image is 3.

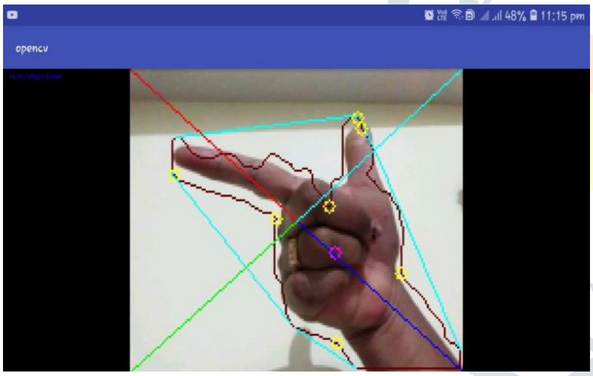


Image representing contour edges of hand

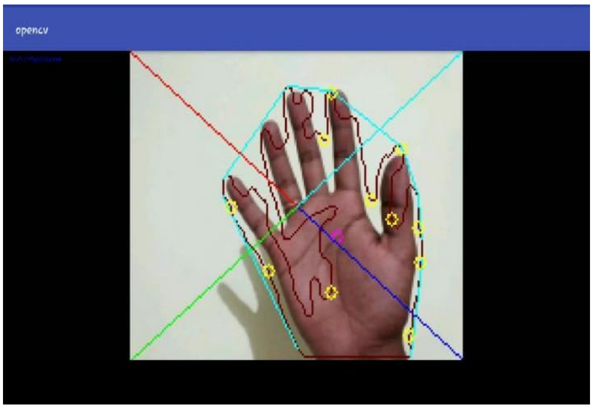


Image representing contour edges of hand

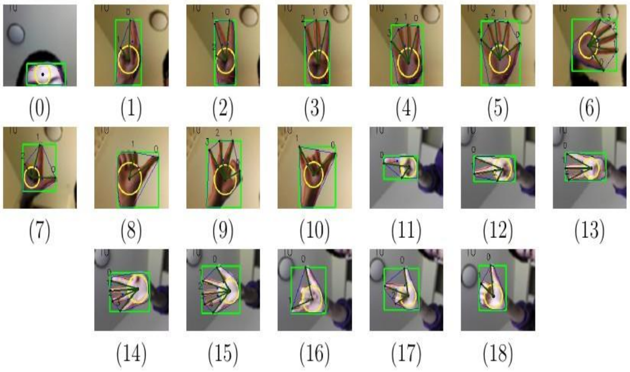


Image representing the hand gestures for numbers and angle calculation between them.

4. Hand Recognition

It is the very next module after hand detection where it is the process where the gesture is recognized by the application and then mapped with the training data present within the database and then an appropriate output is been delivered. To reduce the complexity of feature extraction for hand gestures, the output image of hand portion extraction process was converted into binary image using a thresholding technique. The image so formed may contain some noises. The appearance of such noises may be due to the atmospheric condition at which the images were taken and also the type of source that is used for capturing an image. Hence to remove the noise, the Gaussian filter was generally used. Also, depending upon the percentage of noise present in the binarized image, it was found that the morphological operator namely image erosion can also be used for remove small sharp unwanted details (i.e., noise) from an image. The extent of noise removal is directly proportional to the extent to which a system can be trained correctly and hence classifies the input hand gestures correctly.

A Gaussian filter is a technique used for noise reduction which is a linear filter. It's usually used to blur the image or to reduce noise. If you use two of them and subtract, you can use them for "unsharp masking" (edge detection). The Gaussian filter alone will blur edges and reduce contrast.

The Median filter is a non-linear filter that is most commonly used as a simple way to reduce noise in an image. Its claim to fame (over Gaussian for noise reduction) is that it removes noise while keeping edges relatively sharp. One advantage a Gaussian filter has over a median filter is that it's faster because multiplying and adding is probably faster than sorting.

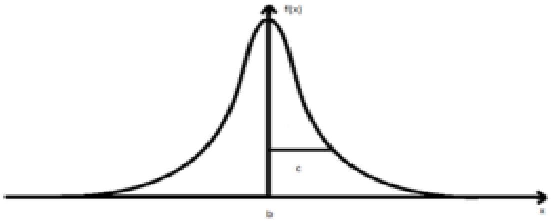
A Gaussian equation is given by

$$f(x) = ae^{-\frac{(x-b)^2}{2c^2}}$$

Here b is the reference point, c is the spread value, d x is distancrom b. As $(x-b)^2$ increases, $e^{(x-b)^2}$ increases and $\frac{1}{e^{(x-b)^2}}$ decreases. Hence, as we move away from reference, the Gaussian value decreases.

When c increases, $\frac{1}{c^2}$ decreases, $e^{\frac{1}{c^2}}$ decreases and $^1/_{e^{\frac{1}{c^2}}}$ increases. Hence, as c increases, the fall as we move away from reference is smaller. Hence the spread is more. A general Gaussian equation is of the following shape:

When c increases, $\frac{1}{c^2}$ decreases, $e^{\frac{1}{c^2}}$ decreases and $^1/_{e^{\frac{1}{c^2}}}$ increases. Hence, as c increases, the fall as we move away from reference is smaller. Hence the spread is more. A general Gaussian equation is of the following shape:



In the above diagram, we can see that the values near the reference point are more significant. This same Gaussian distribution is achieved in a 2-dimensional kernel with the reference point being matrix center. As opposed to Normalized Box filter which gives equal weight to all neighboring pixels, a Gaussian kernel gives more weight to pixels near the current pixel and much lesser weight to distant pixels when calculating sum. The next immediate process after the removal of noise is the feature extraction process, in which, the different techniques are applied based on the type of feature to be extracted. There are various different kinds of distinguished features that can be extracted from the filtered image, but the paper focuses only on the active and in-active finger which is represented by 1 and 0 respectively. So to identify the active and in-active finger, the first task was to identify the finger tip.

4.1 Feature Extraction

The next stage in hand recognition is feature extraction the scale invariant feature transform is used to derive important features from the segmented region. The method retrieves the feature according to the relative position because it does not change from one image to another image. This project basically deals with the design of a system that acquires a user's hand gesture and classifies it based on the predefined hand gestures, stored in a database. The design of a system is basically divided into parts, namely, pre-processing and classification phase. The figures 1 shown below are the list of gestures that the system will recognize it correctly:

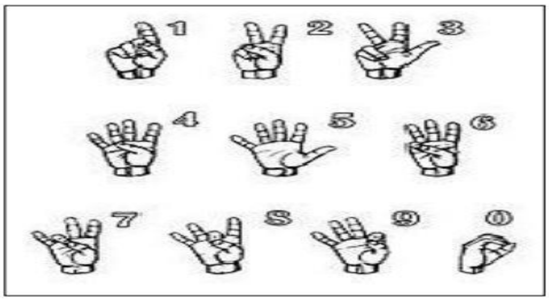


Fig. 1 Gesture representing Numbers

The system will include low-resolution web cam for capturing the hand gestures and an algorithm that processes the acquired images and then classifies the hand gesture correctly. The work mainly emphasizes on the feature extraction from the hand gestures and use that features in the recognition algorithms. Initially, the system will contain a setup procedure, in which, the algorithm is trained on given training set, based on significant feature extracted for different hand gestures. Once the setup in completed successfully, the system will be able to classify the given hand gesture based on the knowledge acquired during the training phase. The design of hand gesture recognition system is broadly divided into two phase. The first phase is the preprocessing phase and the second phase is the classification phase. The efficiency of the Classification phase entirely depends on the preprocessing phase i.e., better the task performed in pre-processing phase, better will be the performance of classification phase. So, all the tasks in pre-processing phase are to be carried out properly. The figures 2 shown below are the list of task for preprocessing phase.

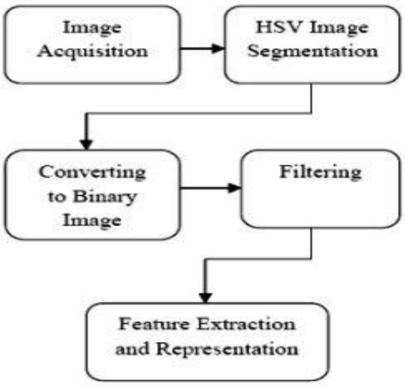


Image Acquisition
HSV Image Segmentation
Converting to Binary Image
Filtering
Feature Extraction and Representation

Fig. 2 Pre-processing Phase

The main purpose of the pre-processing stage is to:
Extract the only hand gesture from an image, Remove the noises (if present) and unwanted region, Process the extracted image to form a binary image and Extract the distinguishable significant features from the processed image, to form a feature set for classification.
Basically, in training phase, the database which consists of list of unique binary pattern for different hand gesture representing some specific number between 0 and 9 and alphabets from A to Z is created and is represented in some form. The Hand gesture classification is done in two phases. The first is the training phase where we train the system with certain predefined symbols and we store them in the database of the system for further classification. The second phase is the testing phase where the system is given the symbols and it compares it with the symbols in the database and compares the most matched feature and gives the output.
There are Gradient, PCA (Principal Component Analysis) and SVM (Support Vector Machine). But, the paper mainly focuses on SVM, which is a machine learning algorithm. Initially, the traditional methods will be followed in preprocessing step, for preparing an image for feature extraction. Once a prepared image (i.e. noise free image), is successfully achieved, the significant features representing a different hand gestures are extracted and represented to include in the classification algorithm. This is because the more features we include in the algorithm the more will be the accuracy of classification.
The SVM is a popular pattern recognition technique with supervised learning. Since it divides the feature space for each class, the SVM can handle unknown data well, although it is not suited to grouping sample data. Support Vector Machines (SVM) is a linear machine with some very good properties. The main idea of a SVM in the context of pattern classification is to construct a hyper plane as the decision surface. The hyper plane is constructed in such a way that the margin of separation between positive and negative examples is maximized. The SVM uses a principled approach based on statistical learning theory know as structural risk minimization, which minimizes an upper bound on the generalization error. The extracted features are stored in database. SVM compares the image with the most matched feature in the database and gives the output on the screen.

4.2 Classification using support vector machine

Classification is an ordered set of related categories used to group data according to its similarities. It consists of codes and descriptors and allows survey responses to be put into meaningful categories in order to produce useful data. It is a useful tool for developing statistical surveys. Classifier is an abstract metaclass which describes (classifies) set of instances having common features. Support Vector Machine (SVM), is one of the best machine learning algorithms, which was proposed in 1990"s by Vapnik. SVM"s are a set of related

supervised learning methods used for classification and regression. A supervised learning is the machine learning task of inferring from supervised training data. A supervised learning algorithm analyzes the training data and produces an inferred function which is called a classifier. SVM has high accuracy, nice theoretical guarantees regarding over fitting, and with an appropriate kernel which can work well even when the data is not linearly separable in the base feature space. The support vector network is a new learning machine for two group classification problems. The machine conceptually implements the following idea: input vectors are non-linearly mapped to a very high dimensional feature space. In this feature space a linear decision surface is constructed. A special property of decision surface ensures high generalisation ability of the learning machine.
The idea behind support vector was previously implemented for the restricted case where the training data can be separated without errors. Support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. A hyperplane is a subspace of one dimension less than its ambient space. A good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class. It is explicitly told to find the best separating line. It searches for the closest points which is called the "support vectors" (the name "support vector machine" is due to the fact that points are like vectors and that the best line "depends on" or is "supported by" the closest points). Once it has found the closest points, the SVM draws a line connecting them. It draws this connecting line by doing vector subtraction (point A - point B). The support vector machine then declares the best separating line to be the line that bisects and is perpendicular to -- the connecting line. The SVM is better because when you get a new sample (new points), you will have already made a line that keeps B and A as far away from each other as possible, and so it is less likely that one will spill over across the line into the other's territory.

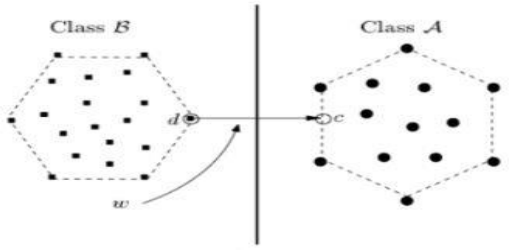
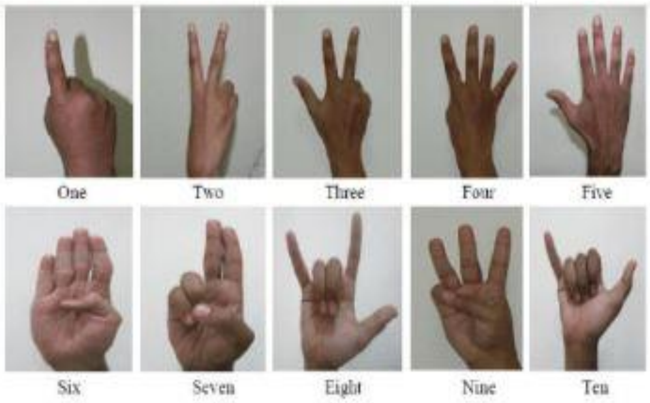


Fig 3 SVM classification.

DATASET FOR CLASSIFICATION



The software analysis is carried out which facilitates matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs. In this section, the performance of real time static hand gesture recognition system in complex background is evaluated for each of the 10 hand gestures. And running input image is used for the performance evaluation. The images were acquired in true colour using a webcam. The images in the testing phase are compared with the database and the result is evaluated.

5. Text to speech translation

Text-To-Speech is a process in which input text is first analysed, then processed and understood Computers do their jobs in three distinct stages called input (where you feed information in, often with a keyboard or mouse), processing (where the computer responds to your input, say, by adding up some numbers you typed in or enhancing the colors on a photo you scanned), and output (where you get to see how the computer has processed your input, typically on a screen or printed out on paper). Speech synthesis is simply a form of output where a computer or other machine reads words to you out loud in a real or simulated voice played through a loudspeaker; the technology is often called text-to-speech (TTS). Talking machines are nothing new—somewhat surprisingly, they date back to the 18th century—but computers that routinely speak to their operators are still extremely uncommon. True, we drive our cars with the help of computerized navigators, engage with computerized switchboards when we phone utility companies, and listen to computerized apologies on railroad stations when our trains are running late. But hardly any of us talk to our computers (with voice recognition) or sit around waiting for them to reply. Professor Stephen Hawking was a truly unique individual—in more ways than one: can you think of any other person famous for talking with a computerized voice? All that may change in future as computer-generated speech becomes less robotic and more human.

IV. RESULTS AND DISCUSSION

This paper is about a system can support the communication between deaf and ordinary people. The aim of the study is to provide a complete dialog without knowing sign language. The program has two parts. Firstly, the voice recognition part uses speech processing methods. It takes the acoustic voice signal and converts it to a digital signal in computer and then show to the user the .gif images as outcome. When the fingers are detected and recognized, the hand gesture can be recognized using a simple rule classifier. In the rule classifier, the hand gesture is predicted according to the number and content of fingers detected. The content of the fingers means what fingers are detected. The rule classifier is very effective and efficient

This application would be designed in the same manner for many other local regional languages. This would help the specially abled people of various parts of the country to learn their regional language. The graphical user interface would be even more enhanced and would be made even more attractive and entertaining. There can be an introduction to local accent and the voice search made by the user could be made to redirect to the sites on World Wide Web. This would help the user to communicate with other people more efficiently. Skin segmentation algorithm can be implemented to extract the skin pixels can have more images of gestures added to the database for the program to recognize captions can be added to the gestures recognized. This would be made available and compatible with IOS such that any version of operating systems would be able to support this application.

REFERENCES

[1] J.P. Bonet. “Reducci_on de las letras y arte para ense~nar a hablar a los mudos”, Coleccion Cl_asicos Pepe. C.E.P.E., 1992.

[2] William C. Stokoe. Sign Language Structure [microform] / William C. Stokoe. Distributed by ERIC Clearinghouse, [Washington, D.C.], 1978.

[3] William C. Stokoe, Dorothy C Casterline, and Carl G Croneberg. “A Dictionary of American Sign Language on Linguistic Principles” Linstok Press, [Silver Spring, Md.], New Edition, 1976.

[4] Code Laboratories. CL NUI Platform. http://codelaboratories.com/ kb/nui

[5] The Robot Operating System (ROS), http://www.ros.org/wiki/ kinect. [6] Open Kinect Project, http://openkinect.org/wiki/Main_Page.

[6] Bridle, J., Deng, L., Picone, J., Richards, H., Ma, J., Kamm, T., Schuster, M., Pike, S., Reagan, R., “An Investigation of Segmental Hidden Dynamic Models of Speech co-articulation for Automatic Speech Recognition.”, Final Report for the 1998 Workshop on Language Engineering, Center for Language and Speech Processing at Johns Hopkins University, pp. 161, 1998.