An Enhanced Gesture Vocaliser for the Vocally Challenged People

K. Deepina Sinthu, G. Kavitha, Prof R.Mariappan
Department of Information Technology
Velammal Institute of Technology
Panchetti. Chennai – 601204
Email deepinasinthu@gmail.com, gkavitha29@gmail.com, mrmrama2004@rediffmail.com

***Abstract*- Gesture recognition is a form of interpreting human gestures through various techniques. "Speech" and "Gestures" are two various forms of communication. In order to provide voice to people who can only communicate through actions, we have proposed a technique to convert the gestures into speech. For this, we have used the technique of image processing. Initially, we store the images of the gestures of dumb people in a database. The input gesture is captured and then the image is processed through various techniques. After processing, the input gesture is compared with the gestures stored in the database and the meaning of that particular gesture is found out. Once the meaning of the input gesture is identified, the next step of voice production is carried out. The meaning of the gesture is produced as voice using speech synthesis. This form of producing voice is similar to the way a human speaks. Thus using our Gesture Vocaliser, we identify the input gesture, process it and find its corresponding meaning. The meaning of the gesture is displayed to help the deaf people and the meaning is also produced as sound to help both the dumb people and blind people. Gesture Vocaliser thus serves as a helping tool for the blind, deaf and dumb people.**

Keywords: Gesture Vocaliser, Image Processing, Neural Networks, Speech Synthesizer

I.INTRODUCTION

Gesture recognition is a topic in computer science and language technology with the goal interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Current focuses in the field include emotion recognition from the face and hand gesture recognition. Many approaches have been made using cameras and computer vision algorithms to interpret sign language. However, the identification and recognition of posture, gait, proxemics, and human behaviors is also the subject of gesture recognition techniques.[1] Gesture recognition can be seen as a way for computers to begin to understand human body language, thus building a richer bridge
between machines and humans than primitive text user interfaces or even GUIs (graphical user interfaces), which still limit the majority of input to keyboard and mouse.

Gesture recognition enables humans to communicate with the machine (HMI) and interact naturally without any mechanical devices. Using the concept of gesture recognition, it is possible to point a finger at the computer screen so that th

cursor will move accordingly. This could potentially make conventional input devices such as mouse, keyboards and even touch – screen redundant.Speech and Gestures are the expressions, which are mostly used in communication between human beings. Research is in progress that aims to integrate gesture as an expression in Human Computer Interaction (HCI). In human communication, the use of gestures and speech is completely coordinated. Machine gesture and sign language recognition is about recognition of gestures and sign language using computers. A number of hardware techniques are used for gathering information about body positioning; typically either image-based (using cameras, moving lights etc) or device based (using instrumented gloves , position trackers etc), although there are some hybrid techniques.

However, getting the data is the first step of the project. The second step, that of recognizing the sign or gesture once it has been captured is much more challenging especially in a continuous stream. A system is developed for recognizing these signs and their conversion into speech. This particular concept is here named as Gesture Vocaliser.

II. GESTURE VOCALISER

Gesture Vocaliser as the name suggests converts the meaning of the gesture into sound or voice. This tool is used to produce voice to assist the dumb people. The additional feature that has been added to our gesture vocaliser the meaning of the gesture is also displayed using Liquid Crystal Display (LCD). This is a portable kit that helps the deaf and dumb people to take it wherever they go.

Gesture Vocaliser is a large scale multi – microcontroller based system being designed to facilitate the communication among the dumb, deaf and blind communities and their communication with the normal people. This system can be dynamically reconfigured to work as a "smart device". Here the gesture vocaliser is implemented by using image processing and voice synthesis. It is of high efficiency and the voice is produced once the gesture is analyzed.

III. EXISTING SYSTEM

Fig: 1 Initial Form Of Gesture Vocaliser

This Gesture Vocaliser was developed for the purpose of converting the gestures of dumb people into text as well as sound. This was accomplished with the help of using a PIC16F877A microcontroller and a Liquid Crystal Display (LCD). It also included another major component which was called as the data glove. The working is explained in the below context.

The vocally challenged people are expected to wear the data glove in either their left hand or right hand. The glove is called as the data glove because each finger is fixed with a sensor. In this case, the sensors used are called as the Flex sensors. The flex sensors provide a change in the resistance value when they are bent in either one of the directions. This is the idea conveyed in this concept. The resistance values of various gestures of dumb people are found out initially and the resistance range is stored as a table of values in the microcontroller. Hence the person wearing the data glove initially starts the conversation with one of his gestures or in a more common way the sign language. Once the gestures are made, there occurs a change in the resistance that flows through the sensors. Based on the obtained resistance value, the values stored in the micro – controller are compared and the correct meaning of the gesture is identified. This identified meaning is the displayed as text using an LCD. The identified meaning is also interpreted as sound. This is done with the help of APR9600.

IV.SYSTEM ARCHITECTURE

The system proposed in this paper is designed to recognize hand gesture in real-time. The technique that is used to recognize hand gesture is bases on computer vision. The overall system architecture is shows in figure 1. The whole system of hand gesture recognition divided into four phases: Image Acquisition, Image Preprocessing, Feature Extraction and Hand Gesture Recognition.

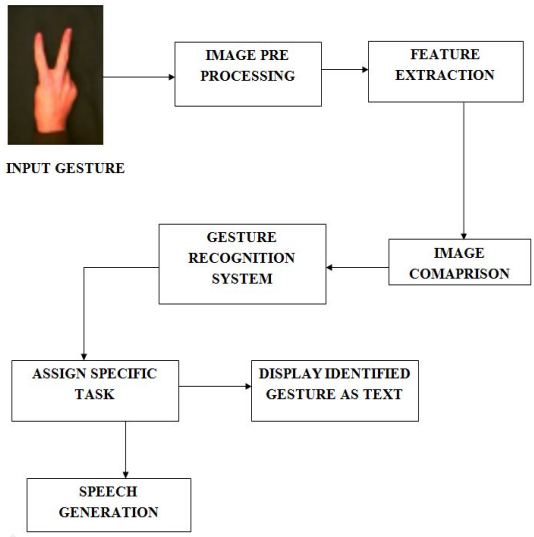


Fig: 2 Architecture of Gesture Recognition System

V.INPUT GESTURE

The gesture that is given as input to the system is called as the input gesture. This input is one of the sign languages of the dumb people.

VI.IMAGE PRE-PROCESSING

The gesture given as input is analyzed. In this pre – processing phase, the image is processed and the features are analyzed.

VII.FEATURE EXTRACTION

The analyzed image is then extracted for its required features. For this particular purpose, six different feature sets of the particular gesture is calculated and only the gesture is extracted for the entire image.

VIII. IMAGE COMPARISON:

After the different features of the particular image are calculated, the extracted image is compared with the images stored in the database. On comparison, the correct gesture is identified.

IX.GESTURE RECOGNITION SYSTEM

Once the gesture is identified in the previous stage, the meaning of the gesture is also found out. The gesture recognition phase has two stages:

X.TEXT DISPLAY

The identified meaning is displayed as text in the display to help in the effective communication of the deaf people.

XI.VOICE PRODUCTION

The next feature here is that the identified meaning is also produced as sound to help in the effective communication of the dumb and blind people.

XII. MODULES

The modules in this project are

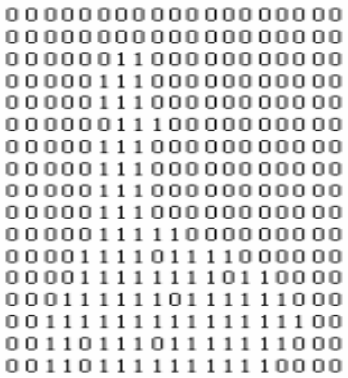
a. Image Processing
i) Image Pre – Processing
ii) Feature Extraction
iii) Image Comparison
b. Voice Generation
i) Voice Recording
ii) Speech Synthesis

XIII. IMAGE PROCESSING

Image Processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image.

A) Image Pre – Processing

Images that are acquiescing from video sequence, passed through the preprocessing phase to obtain enrich data for feature extraction. The algorithms sequentially used in this phase are graying, normalizing and histogram equalizing. After obtaining a grayscale image, we then threshold the image to a binary one. We use the local adaptive threshold algorithm for the binarization step. Then remove the unwanted portion of the image; we have to remove the unnecessary pixels (0) from original image. This is done because we need to develop size independent algorithm.


(a)

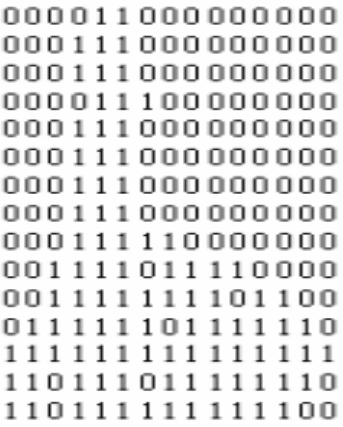

(b)

Fig: 3(a) Original Image (b) Reduced Image

Algorithm

1) Start from top-left corner; repeat for each column and row.
If sum of all black pixels in row/column>0
Then save column and row
2) Else don't save column/row.

B) Feature Extraction

To extract the feature of the hand gesture we used the independent feature extraction method and obtain 6 features for hand gesture.

Center of the image

Center of the image can obtain by using following equation:

x = width / 2… …. … … …. … (1)
y = height / 2... … ….. … … …..(2)

Feature1

The first feature is the relation between the height and the width of the hand gesture

feature 1 = height / width… …. ……(3)

Feature 2 - 5

These features check how the black pixels are distributed in the image. First the number of pixels inside the image is calculated that is total_pixels of hand gestures.

Total_pixel = height × weight…. ….. . …. (4)

The feature 2 and 3 are the percentage of black pixels located in the upper and lower areas of hand gestures, in other words, the pixels located up and down the central point.

feature2 = up_pixels/ total_pixels ……. … (5)
feature3 = down_pixels/ total_pixels…….(6)

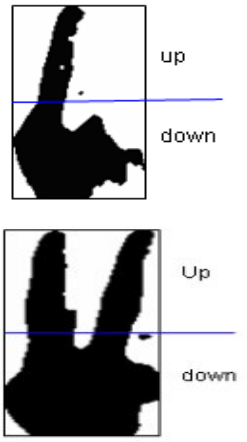


Fig:4 Divide The Image Into Upper And Lower Areas.

The feature 4 and 5 are the percentage of black pixels located in the left and right areas of hand gestures, in other words, the pixels located in the left and right of the central point.

feature4 = left_pixels/ total_pixels…. …..…(7)

feature5 = right_pixels/ total_pixels……... (8)

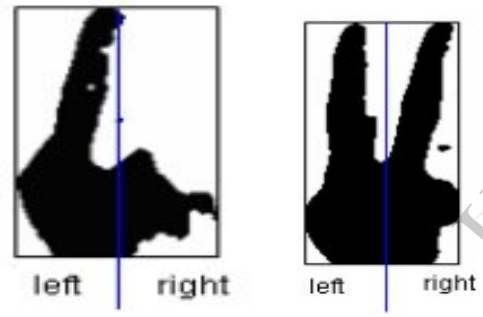


Fig:4 Divide Images Into Left And Right Areas.

Feature 6

The feature 6 calculate average distance between all the black pixels and the central point that invariant object rotation.

Feature 6=

$$(1/(\text{total\_pixel})) \times \sum_{i=0}^{255} \sum_{j=0}^{255} \sqrt{(i-x)^2 - (j-y)^2} \ldots (9)$$

Where (i, j) are the coordinates of a pixel and (x, y) are the coordinates of central point.

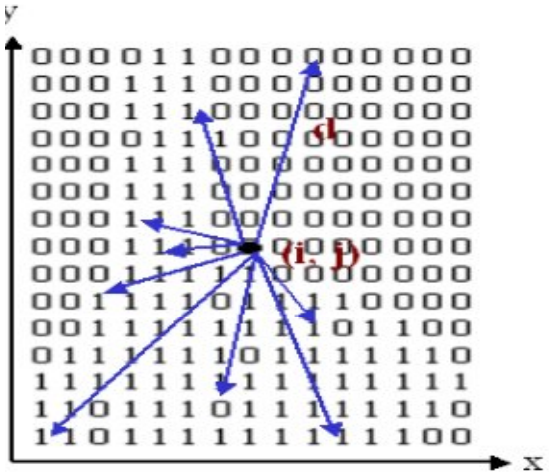


Fig:5 Calculate Distance Between Black Pixel And Central Point

C) Image Comparison

Train applies the inputs to the new network, calculates the outputs, compares them to the associated targets, and calculates a mean square error. If the error goal is met, or if the maximum number of epochs is reached, the training is stopped and train returns the new network and a training record. Otherwise train goes through another epoch. At the beginning the Sum squared error (sse) is big, but after 100 epochs the error of the network is very small and the neuronal net can recognize the training data with a very small error. If the train and test set is same the recognition rates are above 97%, otherwise the results are above 86 %.

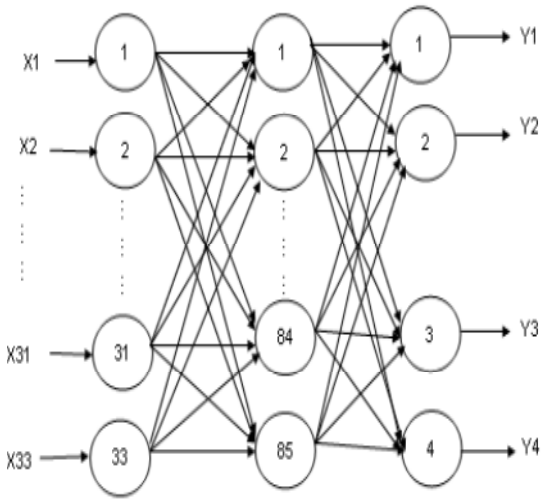


Fig: 6 Network Design For The System

We use 10 frames for each type of hand gesture and evaluate correct recognition rate and the error rate of the system. We use the following equation to find the correct recognition rate:

correct recognition% = correct_recognition / total_frames…….. (10)

And fin the error rate by using following equation:

Error rate = false_recognition / total_frames… … … … … … …. ..(11)

On comparing the feature set (feature 1, feature 2,…… feature 6) values of various gesture, the correct gesture and its meaning is found here in this phase. After this, the sound is produced by the voice generation module.

XIV. VOICE GENERATION

The processed image is then converted to sound in this module. The working is clearly explained below.

A) Voice Recording

Sound recording is an electrical or mechanical inscription of sound waves, such as spoken voice, singing, instrumental music, or sound effects. The two main classes of sound recording technology are analog recording and digital recording. Acoustic analog recording is achieved by a small microphone diaphragm that can detect changes in atmospheric pressure (acoustic sound waves) and record them as a graphic representation of the sound waves on a medium such as a phonograph (in which a stylus senses grooves on a record). In magnetic tape recording, the sound waves vibrate the microphone diaphragm and are converted into a varying electric current, which is then converted to a varying magnetic field by an electromagnet, which makes a representation of the sound as magnetized areas on a plastic tape with a magnetic coating on it.

Digital recording converts the analog sound signal picked up by the microphone to a digital form by a process of digitization, allowing it to be stored and transmitted by a wider variety of media. Digital recording stores audio as a series of binary numbers representing samples of the amplitude of the audio signal at equal time intervals, at a sample rate high enough to convey all sounds capable of being heard.

Digital recordings are considered higher quality than analog recordings not necessarily because they have

higher fidelity (wider frequency response or dynamic range), but because the digital format can prevent much loss of quality fo und in analog recording due to noise and electromagnetic in terference in playback, and mechanical deterioration or damage to the storage medium. A digital audio signal must be reconverted to analog form during playback before it is applied to a loudspeaker or earphones.

B) Speech Synthesis

We are using additive synthesis to synthesize the sound from matrix having rows as different frequencies and columns as time intervals.

C) Additive Synthesis

Additive synthesis is a sound synthesis technique that creates timbre by adding sine waves together. In music, timbre also known as tone color or tone quality from psychoacoustics (i.e. scientific study of sound perception) , is the quality of a musical note or sound or tone that distinguishes different types of sound production, such as voices and musical instruments, string instruments, wind instruments, and percussion instruments Additive synthesis generates sound by adding the output of multiple sine wave generators. Harmonic additive synthesis is closely related to the concept of a Fourier series which is a way of expressing a periodic function as the sum of sinusoidal functions with frequencies equal to integer multiples of a common fundamental frequency. These sinusoids are called harmonics, overtones, or generally, partials. In general, a Fourier series contains an infinite number of sinusoidal components, with no upper limit to the frequency of the sinusoidal functions and includes a DC component (one with frequency of 0 Hz). Frequencies outside of the human audible range can be omitted in additive synthesis. As a result only a finite number of sinusoidal terms with frequencies that lie within the audible range are modeled in additive synthesis.

The sinusoidal waves that are combined above are used to produce sound. The sound produced depends on the features of the wave. The sound can be made audible by use of speakers or ear phones. Thus, the input gestures are converted into voice.

By using display function in matlab, the corresponding meaning of the input gesture is displayed. This is used to help the deaf people.

XV.CONCLUSION

In this project, we have proposed and developed an application to convert the sign language or gestures of dumb people into sound as well as display it as text. The proposed approach applies computer vision methodology that is flexible, sound act in real time performance. This technique

describes a new method to extract gesture features, which make the system rotation, scaling and translation independent. The gesture recognition system recognizes 88.7% hand gestures among the acquiesced frames. In the feature, the system focuses on the three areas. They are-the system can detect and extract human hand from complex image that is an image where a human body is appeared, modify the system so that it can work in any lighting condition and expand the system to recognize the hand tracking.

The sound produced is also very clear so that all the noise from the external source is removed. The voice output produced is hence very clear. The meaning of the gesture is also displayed as text. Hence our project not only helps the dumb people, but also helps the deaf people as well as the blind people.

REFERENCES

[1] A Neural Network Based Real Time Hand Gesture Recognition System in the International Journal of Computer Applications (0975 – 8887) Volume 59– No.4, December 2012 by Tasnuva Ahmed

[2] Analysis and Synthesis of Speech Using Matlab in the International Journal of Advancements in Research & Technology, Volume 2, Issue 5, M ay-2013 373 ISSN 2278-7763 by Vishv Mohan

[3] Gesture Vocaliser Using Data Glove For Vocally Challenged People in the Indian Standard of Technical Education 2013 by K. Deepina Sinthu, G. Kavitha, K. Ranjitha, S. Sangeetha

[4] Face and Hand Gesture Recognition Using Hybrid Classifiers Srinivas Gutta,Jeffrey Huang,Ibrahim F.Imam, and Harry Wechsler

[5] Adaptive Visual Gesture Recognition For Human-Robot Interaction Mohammad Hasanuzzaman,Saifuddin Mohammad Tareeq,Taohang,Vuthichai mpornaramveh,Hironobu Gotoda,Yoshiaki Shirai,Haruki Ueno

[6] Microcontroller Based Gesture Recognition System For The Handicap People Bhavina Patel,Vandana Shah,Ravindra Kshirsagar

[7] Face and Gesture Recognition for Human-Robot Interaction Dr.Md.Hasanuzzaman and Dr.Haruki Ueno

[8] Gesture Spotting and Recognition for Human-Robot Interaction Hee-Deok Yang,A-Yeon Park and Seong-Whan Lee,Senior Member,IEEE

[9]Hand Gesture Recognition System Using Image Processing Abang Irfan Bin Abang Abdul Halil