International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 06 | June-2016 www.irjet.net p-ISSN: 2395-0072

VIRTUAL TALKING SYSTEM WITHOUT SENSOR-IMAGE PROCESSING APPLICATION FOR HUMAN MACHINE INTERFACE TECHNOLOGY

Megha C[1], Dr. Nataraj K R

[1]PG Student, Dept of ECE, SJB Institute of Technology, Banglore, India
[2] HOD and Professor, Dept of ECE, SJB Institute of Technology, Banglore, India

---

Abstract - *Abstract-Every day we see many people who are facing illness like deaf, dumb and blind etc. They face difficulty to interact with others. Previously developed techniques are all sensor based and they didn't gave the general solution. This paper explains a new technique of virtual talking without sensors. An image processing technique called Histogram of gradient (HOG) along with artificial neural network (ANN) has been used to train the image. Web Camera is used to take the image of different gestures and that will be used as input to the Mat lab. The software will recognizes the image and identifies the corresponding voice output which is played using voice replay kit. This paper explains two way communication between the deaf, dumb and normal people.*

Key Words: *Histogram of gradient, virtual talking, artificial neural network, voice replay kit.*

1. INTRODUCTION

India is a country having diversity in culture and also in languages. Because of this People faces difficulty in communication. Then how about the disables as it is difficult for normal people. The deaf and dumb people faces many problems and can be broadly classified into a different categories like interaction with society, language and communication, education, problems in behavior, mental sickness, and safety concerns. Their problems have been worsened because of the absence of a proper technology for them to interact with others.

The technology called virtual talking system uses the flex sensor. Which were attached to the fingers. The flex sensor varies the resistance value because of the stress on fingers. Based on the change of stress on fingers different outputs are given out from sensors. This output will be converted to the digital value and will provide a voice output accordingly. Which will be then played using voice replay kit [1]. The problem here is, flex sensor fixing and the stipulation of sensor output. The proposed paper explains the implementation of the virtual talking without using sensors along with to and fro conversation between the users. This project is an attempt to a sensor less virtual talking machine for Deaf and Mute people. The image processing technique called Histogram of Gradient (HOG) and Artificial Neural network has been used for converting the sign language in to voice output. This project uses camera with the PC and MATLAB installed in it. And it also implements two way communication between the deaf and dumb people and normal people who doesn't know sign languages.

Which means the proposed system is capable of converting the sign language to voice [fig.1], and can also convert the voice input to the sign language output [fig. 2]. Will these people take this system everywhere? No, but they could surely take smart phones.

This is a prototype to develop the concept of converting the sign language to speech. The aim of this paper is to provide an application to the society to establish the ease of communication between the deaf and mute people by making use of image processing algorithm.

2. PROPOSED METHODOLOGY

The steps involved in converting the sign language to voice and voice to sign language is shown in the block diagrams below.

2.1. Block diagrams

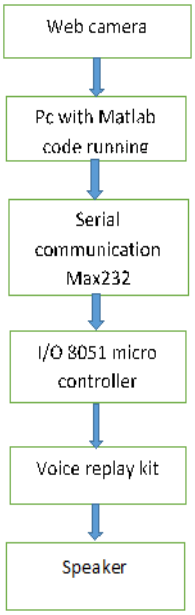

Fig 1. Sign language to voice conversion

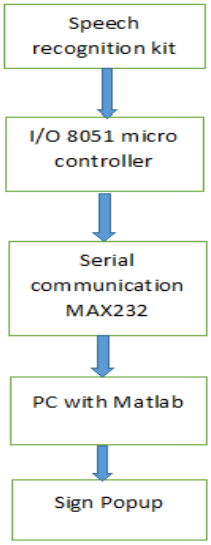

Fig 2. Voice to sign language conversion

The fig3 represents the way of conversion of image captured, which indicates the sign language into a voice output with the help of speaker and voice replay kit. Before identifying the voice corresponds to a particular sign, the system has to be trained with our voice and image for many number of times.

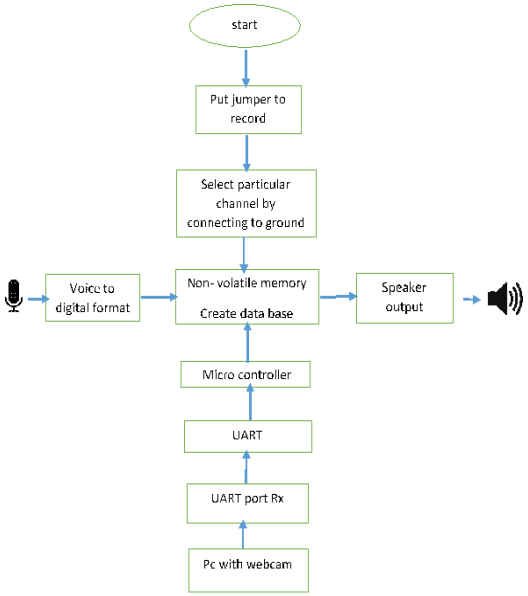

Fig3. Audio playback system

Fig 4 represents the way of conversion of voice received by mic into corresponding image showing sign. This also involves training of device multiple times to select the corresponding image for the received voice.

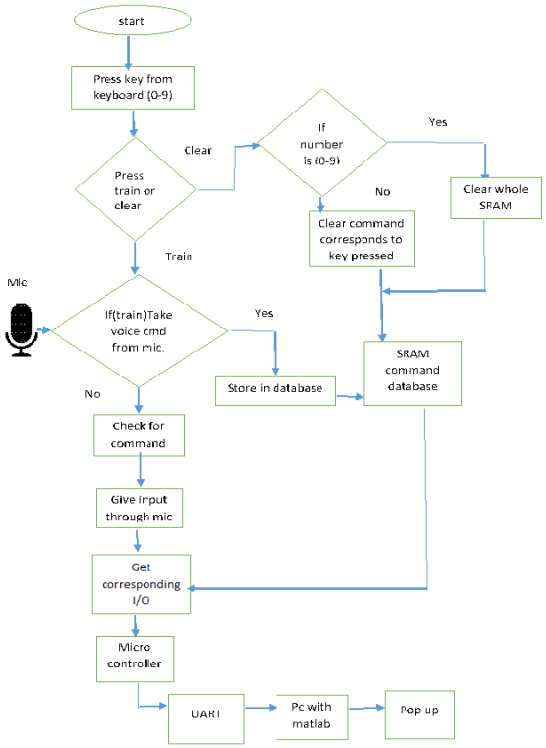

Fig 4. Speech recognition system

2.2. Histogram of gradient(HOG)

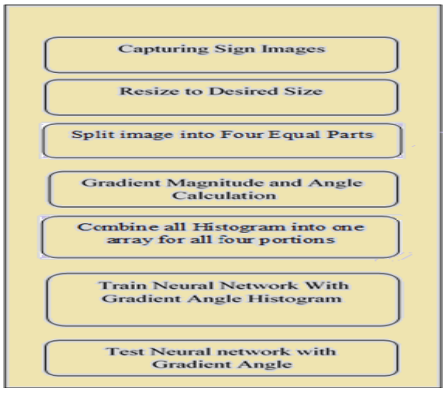

Fig 5. Software implementation steps

HOG descriptors, are descriptors of feature, used in image processing and computer vision .used for purpose of object detection. This technique involves counting the occurrences of gradient orientation in localized portions of the image. These features are then used for the classification of image and for the object recognition. The importance of this HOG is finding the local object appearance and shape within an image and can be described by the distribution of intensity gradients or edge directions. This can be implemented by dividing the image into cells and finding the histogram of gradient direction or edge orientation for the pixels of cell. Then combine the histograms of pixels within the cell. This represents the descriptor.

Instead of using sensors which limits the number of signs, sensor less approach is applied. The sensor less approach needs a vision based approach. The different steps involved in the software are shown in the fig [5]. First step is to capture the image of hand gesture using the camera and this is used as an input to image processing algorithm. The "imresize" inbuilt function is used in MATLAB to vary the pixels size based on the requirement. The next step is to split the image into 4 parts and processed individually using algorithm. Then next step is finding the gradient magnitude and gradient direction of the image. Gradient of an image means directional change in the intensity, using which we can get the information about the image. Histogram of all four portions will be combined to form array which is a descriptor. Then the neural network will be trained using gradient angle of the histogram and will be tested using gradient angle.

2.3. Algorithm steps

- Make the database of sign images
- Get the reference image from camera
- Fink function
- Neural network

a. Image database

With the help of camera we capture the images of the different sign language and those corresponds to the sign language alphabets. The alphabet showing hand pictures will be taken in 10 different directions with slight variation. For example, if we take 10 alphabets as 10 signs (A,B,C,D,E,F,G,H,I,J). We need to capture 100 images. Because each of these alphabets will be captured in 10 different directions with slight variation. Make the database for each alphabet letters using corresponding images. So after these steps we will be having 10 database with 100 images. These database images are applied to the fink function.

b. Reference image

Capture the sign language showing hand image by using webcam. This will be the Reference image. The Reference image will then be used to compare with the database images to find the corresponding voice output.

c. Fink function

The fink function equally divide the image as four portion in column wise direction. Fink function provides direction of each pixel in the image by using ingredient function(predefined MATLAB function).So we get direction for four portions. These direction ranges are -180 degree to +180 degree. We then have to change -180 degree angle into +180 degree by adding +180 degree with negative angles. Now all pixels direction values are changed into positive direction pixel values. Then we equally divide the 180 degree angle into 9 parts. So it gives each part with angle of 20 degree range. The 9 part ranges are,

- Part1 = 0 degree to 20 degree
- Part2 = 21 degree to 40 degree
- Part3 = 41 degree to 60 degree
- Part4 = 61 degree to 80 degree
- Part5 = 81 degree to 100 degree
- Part6 = 101 degree to 120 degree
- Part7 = 121 degree to 140 degree
- Part8 = 141 degree to 160 degree
- Part9 = 161 degree to 180 degree

These values will be then stored to 9 different variables from bin1 to bin9initially zero value will be assigned to all these 9 variables. If a pixel direction values is between the ranges of 0 to 20 degree we will increment bin1 variable value as one. Again if the second pixel direction is between the ranges of 0 to 20 degree, then again we increment bin1 variable value as two. Like this, we increment bin variables depending upon on angle range. Same procedure is applied for four portion. Each portion output contains (1*9) elements. For four portion we get (1*36) elements. This values are stored in one variable (variable name-binfull).This variable is the Fink function output. Same Fink operation is applied for Reference Image also.

d. Neural network

The database output portion of the Fink is applied as an input for neural network. Set targets for each input and choose how many networks we need in neural network and then train the network.

Net=train(net, input, targets)

Then reference image output portion of the fink is applied to trained network Net. It gives output as target which that we give target as training period. By using target value, we can able to find the reference image matched with at which database. For high efficiency we make the same operation as 10 times and then we choose best output from that

3. RESULTS

Fig 6. Image showing match between the test image and database image (A)

Fig 7. Image showing match between the test image and database image (B)

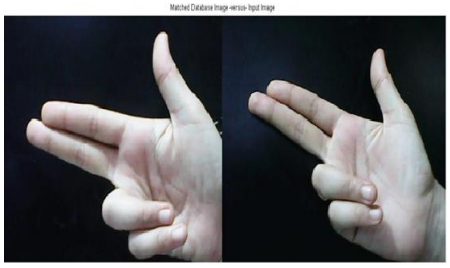

Fig 8. Image showing the match between test image and database image(C)

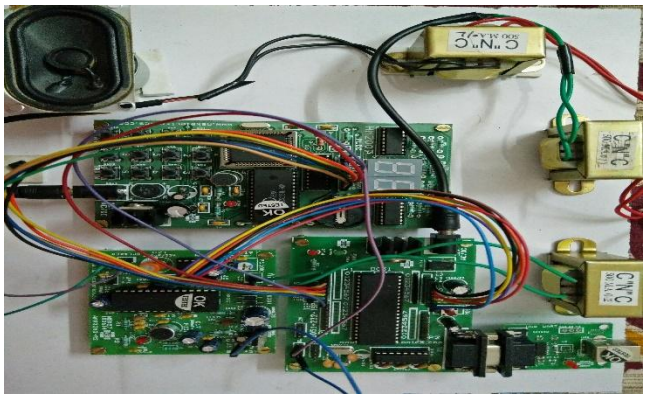

Fig 9 : Experimental setup

4. CONCLUSION

The sensor less Virtual Talk Module is a handy module which provides an easy and satisfactory user communication for deaf and dumb people. The module provides two way communication which helps in easy interaction between the normal people and disables. This module provides a foundation for developing a more durable module to communicate as sentences.

5. STRENGTH

The cost of the module is very negligible compared to its usage and application. The device is low powered since the controlling part is employed using low power. It is possible to implement the device with large set of words.

6. LIMITATIONS

The device communicate only using intended words, it is not communicating in the form of sentences. Proposed system doesn't support for the facial expression. It takes few seconds for the next hand sign once the hand signs are made. Hence the proposed device is slow compared to normal communication. Before using the device it has to be tuned. By using both hands we can increase the number of words and even can create some small statements. Above approach is implemented by capturing the images as snaps from the web camera. Instead of this taking the video for the implementation can be considered. The hardware implementation can be done on the video development boards using DSP processors. The app can be developed for easy handling facilities, so that the disables can carry it easily in a mobile and communicate. This can also be implemented using SIFT match and point pattern algorithms.

REFERENCES

[1]. Vikram Sharma M, Virtual Talk for deaf, mute, blind and normal humans, Texas instruments India Educator's conference,2013.

[2]. Concept and Extent of disability in India from Census 2011 and 2001.

[3]. Indian Sign Language Research and Training Centre. A boon for the deaf and dumb in India

[4]. Q. Munib, et al, American Sign language (ASL) recognition based on Hough Transform and neural Networks Expert Systems with Applications, 2007.

[5]. P.S Rajan and Balakrishnan G, Real time Indian Sign language recognition system to aid deaf dumb people, IEEE 13th International Conference on Communication Technologies, 2011, pp 737-742.

[6]. J. Kim et.al,Bi-channel sensor fusion for an automatic sign language recognition in the 8th IEEE International Conference-2008