Embedded Sign Language Interpreter System For Deaf and Dumb People

Geethu G Nath
Department of Electronics and Communication
Sree Buddha College of Engineering, Pattoor
Kerala, India
geethugnath58@gmail.com

Anu V S
Department of Electronics and Communication
Sree Buddha College of Engineering, Pattoor
Kerala, India
anuvs.ec@gmail.com

Abstract— **Deaf and dumb conversation is hinged on sign language which renders words through hand and finger symbols. Main objective of the project is building a device that assist deaf and dumb people to convey their messages to normal people. The system consists of two main parts. The first part is hand sign recognition and the second one is speech recognition. This ensures a bi directional communication avoiding the assistance of an interpreter. A webcam for image recognition, a microcontroller unit, and a display device associated with the controller and a Bluetooth module constitutes the hand sign recognition part. In the speech recognition part, digitized signal after processing undergoes statistical modeling and matching for recognition. Speech recognition part is integrated on an easy to use android application. Since the system is deployed in both mobile and embedded platforms, a high performance and power efficient ARM Cortex A8 Processor board is used. For image classification Haar classifier is used while 1-D HMM is used for Speech conversion. Sign language recognition has established its importance in many areas such as Human Computer Interactions (HCIs), robotic controls, home automation etc.**

Keywords— **HMM; HCIs; Hand Sign Recognition; Speech Recognition**

I. INTRODUCTION

The paper presents a novel approach of recognizing hand sign for deaf mute people . The system designed ensures a bidirectional communication between the disabled and the normal people. Sign language or finger alphabets are used by those who are earless and speechless . Also it us used by people who caar but cannot talk. Normal people including many of us do not know what exactly the sign language expresses. Hence we find difficulty in communicating with such people. The system proposed in this paper eliminates this barrier of communication. The system is constituted by two main parts. One is hand sign recognition part and the other is speech recognition part. The sign language recognition part is used by deafened and speechless people to channel their ideas to normal people in an audible form The speech recognition part is used by listener to convey their reply to disabled people in visual form.

The sign can be static poses or dynamic gestures depending on whether the sign is movable or not. Static are those hand poses without any displacement where as dynamic signs are string of static poses. The sign recognition part recognizes the signs by comparing it with database created in the training phase. Training of the system is done usingHaar training method. During the training phase hand gestures are fed to the database by taking a number of samples of each sign and the probability of occurrence of each sample is also provided. In speech recognition system the audio reply of the listener is reformed to a visible form so that the deaf and dumb people are able to see it. The audio is transformed to text format using an application developed in smart phone.

The hand sign recognition part is implemented in an ARM CORTEX processor board. The accelerated processing promotes real time performance. This paper is centered around dynamic sign recognition and Beagle Bone Black is used as the mini central processing unit embedded with a debian Linux operating system supported by Open Computer Vision Library (Open CV). Various process involved are image acquisition, skin color recognition, hand segmentation and feature extraction. A USB camera is utilized as sensing device for image acquisition and preprocessing steps are performed by the ARM CORTEX A8 processor board. The software tool Open CV provides necessary supporting functions for image processing. To extract features of the acquired image the centroid of the image is calculated. The finger count is obtained by determining the Euclidean distance from the centroid to finger tips, silhouette of the hand image, the convexity points and defect points are obtained for recognition. Open CV is a powerful software tool that reduces the complexity of these image processing steps. [12] Different classification techniques are used to categorize the image and the corresponding text is obtained. This is read by the the normal people in their smart phone.

978-1-5090-3294-5/17/$31.00 ©2017 IEEE

II. RELATED WORKS

Hand sign recognition is the area of research for many years. There are different methods for hand sign recognition employing different techniques.[1] Approaches that can recognize hand gestures are of two types- data glove based method and vision based method. Sensor devices are utilized for digitizing hand movements into data in data glove method .These measurements are accurate and fast and are very expensive. Vision based methods requires a camera to achieve natural interaction without use of extra devices. The vision based methods can be of two categories- 3-D model based and appearance model based.

3D model based method depends on 3-D kinematic hand models. The input image and 2-D appearance of the hand models are compared [6]. This determines the hand postures. In appearance based models extract image features are extricated and the appearance of hand images are modeled. This method is prevalent over other techniques for real time processing. Segmenting out the hand regions from background is the first step. Skin colored regions from the images is identified to segment background [7]. This method is popular, but is responsive to lighting conditions and no other skin-like objects should be in the images. Other often used features are Haar-like features were proposed to recognize human faces. Hand gesture classifiers proposed were trained with Adaboost and Haar like features to improve the accuracy of classification. [14]. The Adaboost learning algorithm improved classification accuracy by adaptively selecting the best features. Hand detection with Adaboost that adopted modified Haar-like features and the detection of scale-space features to find palm and finger-like structures.

Many gesture-recognition systems use the help of inexpensive color-coded gloves for hand segmentation, which simplify the filtering out of other objects and backgrounds. Such approaches have been tested and approved by many researchers. One of the systems used a glove with six colors. A particular color was used to cover each finger and palm. Some other techniques engraved patterns in a normal cloth glove that decipher the evaluation of hand symbols.

After segmentation the next process to be performed is feature extraction and classification that can be done using different mapping techniques. One shape matching that rely on contour that make use of Fourier descriptors and wavelet descriptors. B- Spline is another shape matching technique based on contour. Another shape matching method is local features based object matching which used SURF and SIFT algorithms [11]. Two classical approaches for extracting the analytical visage and trimming the volume are Principal Component analysis (PCA) and Linear Projection Method (LPP) [13]. Hand sign recognition based on PCA and LPP to extract feature vectors after the hand shape samples are normalized in rotation and size. Another system uses finger and palm arrangements for hand sign recognition [12]. It employs contour points and convex hull or Jarvis algorithm to condense the features for posture recognition. Rectilinear and unbend signs can be recognized by this technique.

In this system we are employing canny edge detection algorithm for edge detection, 2-D HMM for matching process for image recognition. For speech recognition 1-D HMM based techniques are used.

III. PROPOSED SYSTEM

The preeminent target of the system is to build an assistance for communication of deafened and speechless people. This waive the fence for communication between normal and disabled people. This eliminates the help of an interpreter for communication. Normal people can also use the device for sign language learning. The reply of the people in audio form can be made visible to the deaf and dumb people using the android application . The device also can be employed in different applications like HCIs, home automation, etc.

A. Hand sign Recognition Module

Architecture employed is shown in Fig.1. From the physical environment image is acquired using a camera module that can be directly connected to the controller board. Subsequent step is preprocessing that include skin color extraction, reducing dimension using blurring, RGB to binary conversion, edge detection. After preprocessing image feature is extracted by different steps like obtaining the contour, calculating the centroid, determining the hull and defect points etc. This gives the corresponding number values by identifying the number of fingers. Finger alphabets can be recognized using matching algorithms from the database created during the training phase.

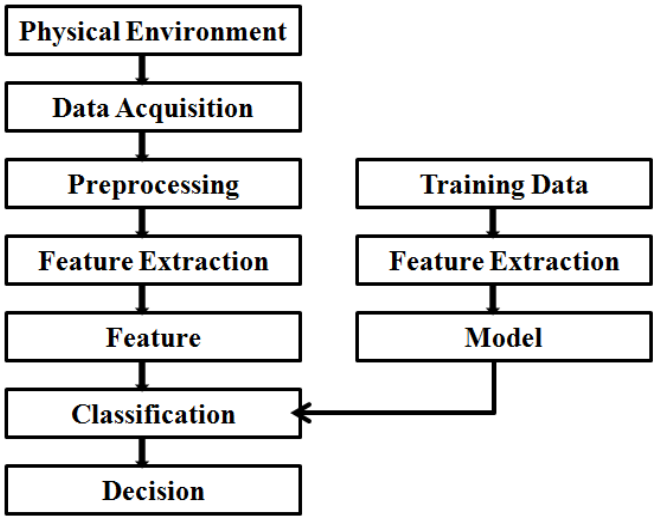

Fig 1. System Architecture

In the preprocessing stage main process happening are skin color extraction, blurring of image to reduce dimension and pixel values , RGB to binary conversion that is thresholding, edge detection etc.

Blurring or smoothing is an image processing operation done as a dimension reduction operation. The logic for

smoothing are many. Smoothing is performed with the intension of reducing noise .We employs a filter to our sign image to accomplish smoothing procedure. The most common type of filters are linear, in which an output pixel's value g(i,j) is determined as a weighted sum of input pixel values f(i+k,j+l).

$$g(i,j) = \sum_{(k.l)} f(i+k,j+l)h(k,l) \quad (1)$$

h(k.l) is called the kernel, which is the coefficients of the filter. The filter is pictured as a window of coefficients that drift across the image.

Thresholding is the process of segmentation by which the zone of concern are isolated from background. This partition is rooted in the fluctuation of intensity values between the object and surroundings pixels. A Threshold value is set and each pixel intensity value is compared with this threshold value. The binary value of 0 (black) and 255 (white) is given to these parted pixels. The thresholding operation can be given as

$$d(x, y) = \begin{cases} \text{max value, if } s(x, y) > \text{thresh} \\ 0, \text{otherwise} \end{cases} \quad (2)$$

d(x, y) is the threshold image and s(x, y) is the source image. After thresholding canny edge detection algorithm is used. Sobel kernel is used to filter the image. Filtering is done in both horizontal and vertical direction. Thus we obtain the first derivative in horizontal direction Kx and vertical direction Ky . From these two images, we can find edge gradient and direction for each pixel as follows:

$$\text{Edge-Gradient (K)} = \sqrt{K^2x + K^2y} \quad (3)$$

$$\text{Angle } (\theta) = \tan\text{-}1(Kx/ Ky) \quad (4)$$

Gradient direction is always at an angle of 90 degree to the edges. Four angles represents the vertical, horizontal and two diagonal directions .Gradient is rounded to one of the four angles. For recognizing numbers, convex hull detection technique or Jarvis algorithm is used A group of points is taken in a plane and a Convex polygon is obtained which constitutes all the points in that plane. This polygon is called convex hull. which is obtained by starting from the point with minimum x coordinate value and covering points in counterclockwise direction. Next point is selected as the point that beats all other points at counterclockwise orientation, i.e., next point is q if for any other point r, then "orientation(p, r, q) = counterclockwise". For identifying the other symbols in sign language, matching algorithm is performed. Initially a database or model is created during training of the system. Input images are fed to the system on training and feature vectors are extracted to create a model . This model is utilized to perform the matching or classification of the input image obtained. Image is recognized after classification and corresponding decision is made.

B. Speech recognition system

Speech recognition in humans is performed by ear and brain.The sound waves reaching the ears are passed to the brain which is the processing part. The block diagram for automatic speech recognition system is shown in Fig 2. The speech signal is converted to text and is shown on the application module. This is then send to the device held by deaf and dumb people using the Bluetooth module.

Speech processing stage includes digitizing, reducing background noise from the signal, identifying variability in individual speech. Approaches to automatic speech recognition s+ystem include template based methods, statistical based methods and neural network based approaches. 1 D Hidden Markov model algorithm is used for speech recognition .

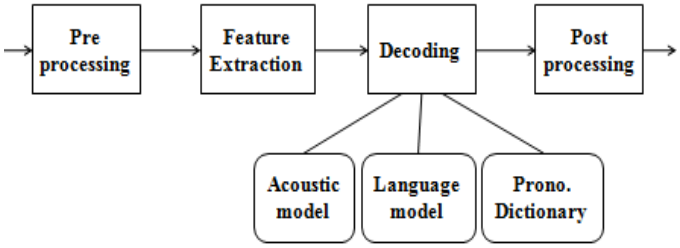

Fig 2. Speech recognition system block diagram

C. Speech Recognition Using HMM

A Speech Recognition System can be operated in 4 main steps including analysis, feature extraction, modeling and testing. Analysis include data regarding the identity of speaker. Extraction partitions speech from one another and find out its features. HMM modeling technique is used for speech recognition. During the training phase words are prerecorded using PCM technique and the file is saved. Linear Predictive Coding method is used to perform speech analysis and synthesis. Speech samples undergo preprocessing techniques, then for each model probability of getting the observation sequence is calculated. The spoken word is recognized based on maximum likelihood estimation. HMM based speech to text (STT) is shown in Fig 3.

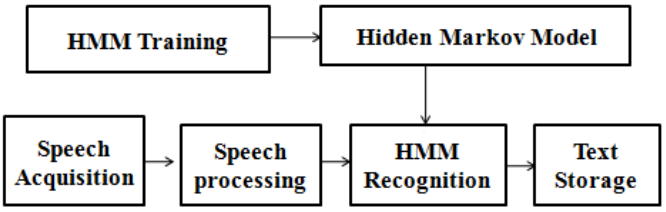

Fig 3. Speech Recognition Architectiure

IV. IMPLEMENTATION

A. Hardware Implementation

A USB camera for image acquisition and processing steps are performed by the ARM CORTEX A8controller board. The hardware architecture is shown in Fig.4.

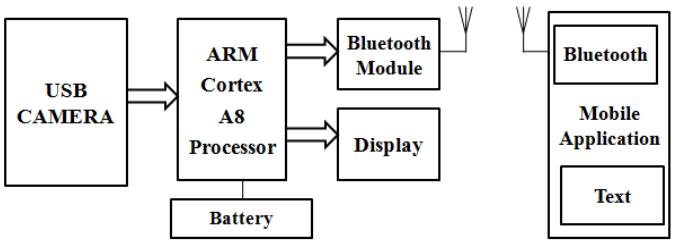

Fig 4. Hardware Architecture

ARM CORTEX A8 a 32 bit microprocessor with 1GHZ clock frequency is used. It supports high level operating system and supports Debian, Android, Ubuntu operating systems. Low power consumption is an added advantage. Display device is used to visualize the text of the corresponding hand image

B. Software tools used

Open CV (Open Source Computer Vision Library) is a library which mainly focuses at real-time computer vision. It supports programming languages such as c, c++ and java. Various operating systems like Windows, Linux, Mac OS, IOS and android are supported by open CV. Open CV has its advantages in computational efficiency in real time applications. The collection of more than 2500 algorithms make this tool suitable for computer vision and machine learning applications. For real rime image processing open CV can provide all the basic data structures.

C is the most powerful language used for embedded applications. As time progressed, use of microprocessor assembly language reduced . Embedded systems go for C as the embedded programming language of choice. C is the most widely used programming language for embedded processors/controllers. Assembly can also be used where we can compromise for properties like high timing accuracy, code size efficiency, etc. Hence C is preferred.

V. EXPERIMENTAL RESULTS

The hand sign recognition system is implemented in a board containing ARM CORTEX A8 processor. The software tool used is Open CV which contain real time image processing capabilities. Open CV supports languages like python, c, c++, and java. The language used is C . Hardware implementation of the sign recognition system is shown in Fig 5.

Fig 5. Hardware Implementation

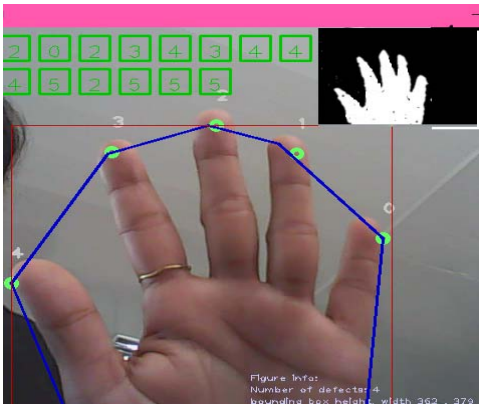

Fig 6. Sign recognition of numbers

The sign language for numbers were recognized and converted to text. This output is shown in fig 6. Words and sentences can be converted to text and send to the mobile device using Bluetooth module and can be read on the developed application. The audio conversion is also performed by the user friendly application. Fig.7. shows the Input and Output of Speech recognition application

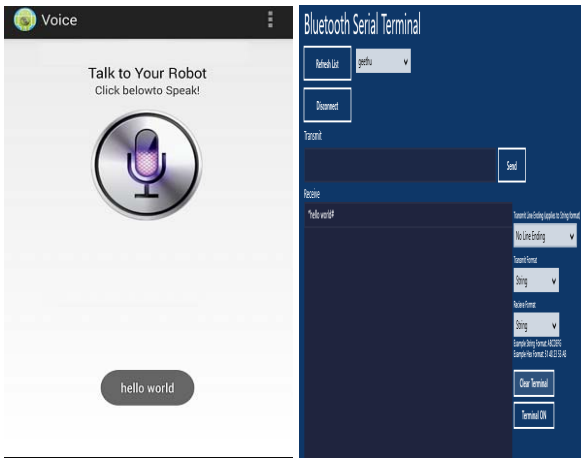

Fig 7. Input and Output of Speech recognition application

VI. CONCLUSION AND FUTURE SCOPE

Hand sign recognition system was implemented using ARM CORTEX development board. The system was developed to recognize sign language used by deafened and speechless people. The system can assist deaf mute people to convey their messages to normal people. This system can be used for a variety of applications including assistance to deaf mute people. Moreover the device can be used as a sign language learning aid. Application developed is capable of converting the audio reply to text and send it to the system in deaf and dumb people. This ensures two way communication between the disabled as well as normal people. This eliminates the need of interpreter in between.

The system along with the application is capable of recognizing both image and speech. Hence it can be used for many applications like gesture controlled robot for physically challenged, gesture controlled doors and vehicles, gesture

controlled keyboard and mouse to interact with computer, gesture controlled appliances like air conditioner. Visually impaired people can use the speech recognition application for different application like automatic call routing, voice controlled devices and other applications.

REFERENCES

[1] Hiroomi Hikawa, Member, IEEE, and Keishi Kaida "Novel FPGA Implementation of Hand Sign Recognition System With SOM–Hebb Classifier" *IEEE Transactions on Circuits And Systems For Video Technology, Vol. 25, No. 1, January 2015*

[2] Su Myat Mon, Hla Myo Tun, "Speech-To-Text Conversion (STT) System Using Hidden Markov Model (HMM)", *International Journal Of Scientific & Technology Research Volume 4, Issue 06, June 2015*

[3] Christoph S. Garbe and Bj¨orn Ommer, "Parameter Estimation in Image Processing and Computer Vision".

[4] Wang F, Ngo C W, Pong T C. " Simulating a smart body by real time gesture detection in lecture videos." *IEEE Transactions on Multimedia, 2008, 10(5):926-935.*

[5] S. Ju, M. Black, S. Minneman, D. Kimber, "Analysis of Gesture and Action in Technical Talks for Video Indexing," in *IEEE Conf. on Computer Vision and Pattern Recognition,CVPR 97*

[6] A. J. Heap and D. C. Hogg, "Towards 3D hand tracking using a deformable model," in Proc. 2nd Int. Conf. Autom. Face Gesture Recognit., Oct. 1996, pp. 140–145.

[7] W. K. Chung, X. Wu, and Y. Xu, "A realtime hand gesture recognition based on haar wavelet representation ," in Proc. IEEE Int. Conf. Robot.Biomimetics, Feb. 2009, pp. 336–341.

[8] M. V. Lamar, M. S. Bhuiyan, and A. Iwata, "Hand gesture recognitionusing T-CombNet: A new neural network model," IEICE Trans. Inform. Syst., vol. E-38-D, no. 11, pp. 1986–1995, 2000.

[9] B. Wang and J. A. Bangham, "PCA based shape descriptors for shape retrieval and the evaluations," in Proc. Int. Conf. Comput. Intell. Security, vol. 2. Nov. 2006, pp. 1401–1406.

[10] N. Gamage, K. Y. Chow, and R. Akmeliawati, "Static hand sign recognition using linear projection method," in Proc. 4th Int. Conf. Auto. Robots Agents (ICARA), Feb. 2009, pp. 403–407.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints,"Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, 2004

[12] 24] G. Tofighi, S. A. Monadjemi, and N. Ghasem-Aghaee, "Rapid handposture recognition using adaptive histogram template of skin and hand edge contour," in Proc. 6th Iranian Mach. Vis. Image Process. (MVIP), Oct. 2010, pp. 1–5.

[13] X. He and P. Niyogi, "Locality preserving projections," in Advances in Neural Information Processing Systems, vol. 16. Cambridge, MA, USA:MIT Press, 2004, pp. 153–160.

[14] Q. Chen, N. Georganas, and E. Petriu, "Real-time vision-based hand gesture recognition using Haar-like features," in Proc. IEEE Instrum. Meas. Technol. Conf. (IMTC), May 2007, pp. 1–6.

[15] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(24):509–521, 2002.

[16] S. Bres and J.-M. Jolion. Detection of interest points for image indexation. In L. N. i. C. S. Springer Verlag, editor, Proceedings of the third international conference on Visual Information and Information Systems, volume 1614, pages 427–434, 1999.